Enshrouding & Star-planet Interactions in Close-orbiting Systems

Thesis

Enshrouding & Star-planet Interactions in Close-orbiting Systems

James Doherty

A thesis submitted for the degree of Doctor of Philosophy

in Astronomy

October 9, 2020

‘What is between the star and the sea?
A bird as bright as a bird can be,
What is between the bird and me?
Only a star, only the sea.’

The Waterboys

Abstract

The OU-SALT survey of the chromospheric activity of close-orbiting, transiting exoplanet hosts is unique. Emission flux in stellar Ca II H & K line cores is measured using the same telescope, spectrograph and calibration. I present homogeneous $\log(R'_{\rm HK})$ values for 104 bright southern hemisphere stars, ~ 90 % of which host hot Jupiters. Errors associated with systematic offsets between instruments are avoided. Uncertainties are on average three times smaller than in prior state-of-the-art planetary research. Approximately one third of the OU-SALT main sequence sub-sample is less active than a quiet F, G or K star that is devoid of spots, faculae and active regions (corresponding to $\log(R'_{\rm HK}) < -5.1$). Only 2 % of field stars display such sub-basal activity. Absorption in the interstellar medium can not account for these depressed activity measurements. This provides compelling evidence of atmospheric escape from highly irradiated planets, and the absorption of emission flux by diffuse enshrouding discs. I find that such discs form only around stars with effective temperatures of 5200 K $\lesssim T_{\rm eff} \lesssim$ 6600 K and masses of $M_\star \gtrsim 0.8\ M_\odot$.

I present evidence of star-planet interactions (SPI) in the OU-SALT sample with $4\,\sigma$ certainty. I compare two multivariate models that predict activity: one that incorporates a proxy term for SPI (M1) and one that does not (M2). M1 consistently outperforms M2, providing a $0.003 - 0.3$ % probability that massive, close-orbiting planets do not influence host activity. A systematic correlation study reveals that $\log(R'_{\rm HK})$ is correlated with planetary surface gravity (Spearman's coefficient $\rho = 0.41 \pm 0.02$), semi-major axis ($\rho = -0.43 \pm 0.01$), and planet mass ($\rho = 0.52 \pm 0.01$), as well as several SPI proxies. This provides further evidence of atmospheric escape and SPI in the OU-SALT sample. I confirm that ultra-hot Jupiter host WASP-43 has anomalously high activity ($\log(R'_{\rm HK}) = -4.18^{+0.09}_{-0.11}$), likely caused by tidal spin-up. Significant activity variation is identified in several systems, notably KELT-11. A novel method to search for orbitally-modulated SPI in planet host samples is formulated.

I search for the spot-modulated rotation period ($P_{\rm rot}$) of terrestrial planet host candidate HD 184960 using the Lomb-Scargle periodogram, string-length method, and phase dispersion minimisation. I constrain the star's photometric amplitude to be less than 0.018 mag and estimate that $P_{\rm rot} \approx 5 - 6$ d.

Acknowledgements

I would like to thank my supervision team: Carole Haswell for creating a fascinating research opportunity; John Barnes for reviewing the chapters of this thesis; and Ulrich Kolb for assistance observing with PIRATE.

I am grateful to the Open University for offering a lawyer a pathway to become an astronomer, and to the Science Technology and Facilities Council for funding my work. Thank you to the astronomers of the Southern African Large Telescope for collecting our OU-SALT survey data, and for answering innumerable queries. Thanks to Fraser Lewis and colleagues at the Las Cumbres Observatory. I extend my gratitude to ESO and OPTICON for arranging unforgettable observing runs under the world's most vivid skies.

Thanks to my collaborators: Dan Staab, my go-to buddy in the early stages; Luca Fossati for useful conversations at EPSC; James Jenkins for his hospitality in Santiago (including an animated afternoon watching the Scottish football team); Joe Cooper for coding assistance; Richard Busutil for assistance with his Roche scripts; and Geoff Bradshaw for helping to solve IT challenges. I would also like to thank Andrew Norton, Marcus Lohr, Simon Green and Suzanne Aigrain for their contributions to my work.

Thanks to my office mates: too many to identify all but honourable mentions to Jimmy for the welcome on Day 0, Jana for lessons in Slovak, Meredith and Dean for help with PIRATE & COAST, Ayo for the nickname, Rachel for the coffee, Heidi for the chats, and Mark Parker for the laughs.

Thanks to my parents, J.G. and Marie, and my sisters, Sinéad and Éadaoin, for sound advice and counsel throughout this process. A warm mention to my good friends Donal and Mie—cheers for the late-night discussions on boosting and wiggling.

My most heartfelt thanks is reserved for my partner Laura, without whose support I could not have completed this journey. I dedicate this thesis to my wee lassies Nuala and Saoirse, both of whom arrived during the course of this project.

Contents

Abstract iii

Acknowledgements v

1 Introduction 1

1.1 A Nobel pursuit . . . 2

1.2 The known exoplanet population . . . 4

1.2.1 Detection . . . 4

1.2.2 Close-orbiting systems . . . 7

Hot Jupiters . . . 7

Ultra-hot Jupiters . . . 8

Super-puffs . . . 9

Ultra-short period planets . . . 9

Disintegrating rocky planets . . . 10

1.2.3 Demographic features . . . 11

The Neptune Desert . . . 11

The Evaporation Valley . . . 12

The Mass-period Gap . . . 13

1.3 Stellar activity . . . 14

1.3.1 Magnetic activity in stars . . . 14

1.3.2 Useful spectral lines . . . 14

1.3.3 Chromospheric activity . . . 15

The Mount Wilson system . . . 16

Basal limit . . . 18

The Vaughan-Preston Gap . . . 19

1.3.4 Stellar rotation and chromospheric activity . . . 19

Rotational modulation and period 20
1.4 Atmospheric escape 21
1.4.1 Observations of atmospheric escape 21
Lyman α 21
Helium triplet 22
UV and X-ray observations 23
1.4.2 Mass loss mechanisms 23
Non-thermal escape 24
Thermal escape 24
1.4.3 Evolution of mass-losing planets 27
Hot giant planets 27
Ultra hot Jupiters 28
Low mass planets 28
1.5 Enshrouded systems 30
1.5.1 WASP-12b 30
1.5.2 The Dispersed Matter Planet Project 34
1.6 Star-planet interactions 35
1.6.1 Mechanisms 35
Tidal SPI 35
Magnetic SPI 36
Timing 38
1.6.2 Observational evidence 38
Individual systems 39
Population studies 41
Lessons in SPI 42
1.7 Thesis overview 43

2 The rotation period of HD 184960 **45**
2.1 HD 184960 46
2.2 Observations 47
2.3 Photometry 48
2.3.1 Quality checks 51

2.3.2 LCO pipeline . . . 52
2.4 Period search . . . 54
2.4.1 Lomb-Scargle periodogram . . . 54
2.4.2 String-length method . . . 58
2.4.3 Phase dispersion minimization . . . 59
2.4.4 Phase-folded light curves . . . 62
2.5 R-band observations . . . 65
2.5.1 Correcting for offsets . . . 65
2.5.2 Period searches . . . 66
2.5.3 Phase-folded light curves . . . 70
2.6 Discussion . . . 72
2.7 Summary . . . 75

3 The OU-SALT survey **77**
3.1 Survey value . . . 78
3.2 Observations . . . 80
3.3 Data reduction . . . 80
3.3.1 Image trimming . . . 82
3.3.2 Wavelength calibration . . . 82
3.3.3 Flat-fielding . . . 84
3.3.4 Cosmic ray removal . . . 84
3.3.5 Background subtraction and spectral extraction . . . 85
3.3.6 Specific challenges . . . 86
3.4 S-value extraction and conversion to $\log(R'_{\rm HK})$ . . . 88
3.5 Homogeneous activity survey . . . 90
3.5.1 Uncertainties . . . 90
3.6 Comparison to published activity values . . . 96
3.6.1 Systematic offset . . . 98
3.7 Summary . . . 99

4 OU-SALT systems **101**
4.1 Target selection . . . 102
4.2 Stellar evolutionary status . . . 103

4.2.1 Distance to target . . . 103
4.2.2 Hertzsprung-Russell diagrams . . . 104
4.2.3 Host classification . . . 106
4.3 Stellar age . . . 115
4.4 Planets . . . 117
4.4.1 Demographics . . . 117
4.4.2 Incident flux . . . 117
4.4.3 Roche lobe filling . . . 120
4.5 Star-planet interaction proxies . . . 122
4.5.1 Tidal star-planet interactions . . . 122
4.5.2 Magnetic star-planet interactions . . . 124
4.6 Temporal activity variations . . . 124
4.6.1 WASP-43 . . . 124
4.6.2 WASP-103 . . . 127
4.6.3 KELT-11 . . . 129
4.6.4 HD 73256 . . . 130
4.6.5 Other systems . . . 130
4.7 A novel search for orbitally-modulated SPI . . . 132
4.8 Summary . . . 137

5 OU-SALT population study **139**
5.1 Scope of study . . . 140
5.2 Activity in context . . . 140
5.3 ISM absorption . . . 146
5.3.1 Correcting for absorption . . . 147
5.4 Planet parameters . . . 152
5.4.1 Surface gravity . . . 152
5.4.2 Equilibrium temperature . . . 157
5.4.3 Semi-major axis . . . 159
5.4.4 Mass . . . 163
5.5 SPI proxies . . . 164
5.6 Stellar parameters . . . 172

5.6.1 Effective temperature . . . 172
5.6.2 Mass . . . 172
5.6.3 Rotation period . . . 174
5.7 Interdependencies between variables . . . 177
5.8 Constraining enshrouded systems . . . 180
5.9 Summary . . . 183

6 Multivariate analysis of the OU-SALT MS sample 185
6.1 Multivariate statistics . . . 186
6.2 Predicting activity . . . 187
6.2.1 Multiple linear regression model . . . 187
Principal component analysis . . . 188
Data scaling . . . 189
Calculating PCs . . . 190
Linear model with PC input variables . . . 191
Estimating prediction uncertainty . . . 194
6.2.2 OLS-PCA results . . . 194
Assessing prediction power . . . 197
Significance . . . 199
6.3 Linear model optimization . . . 201
6.3.1 Gaia DR2 data . . . 201
6.3.2 Non-linear PCA . . . 202
6.3.3 Predictor variable optimization . . . 204
6.4 OLS-kPCA model . . . 205
6.4.1 Results . . . 207
6.5 Sub-sample tests . . . 213
6.6 Predictor variable tests . . . 214
6.7 Summary . . . 218

7 Conclusions 221
7.1 Thesis summary . . . 221
7.1.1 Rotation period of HD 184960 . . . 221
7.1.2 The OU-SALT survey . . . 223

7.2 Limitations . . . 231
7.3 Future work . . . 232

A Principal components analysis & model success **235**
A.1 Linear PCA . . . 235
A.2 Kernel PCA . . . 237
A.3 Types of kernel . . . 240
A.4 Assessing model success . . . 241
A.4.1 Residual-based analysis . . . 241
A.4.2 Likelihood-based metrics . . . 242

B Supervised learning approach **243**
B.1 Introduction . . . 243
B.2 *XGBoost*-kPCA model . . . 245
B.2.1 Predictor variable optimization . . . 245
B.2.2 Tuning *XGBoost* . . . 246
B.2.3 Cross validation . . . 247
B.3 Results . . . 248
B.4 Summary . . . 250

Bibliography **253**

Chapter 1

Introduction

In this opening chapter, I explore and discuss the main themes of my thesis. I begin with an introduction to exoplanetology in Section 1.1, before providing an overview of the planet population detected to date in Section 1.2. The magnetic activity of convective stars—a phenomenon central to this work—is discussed in Section 1.3. Observations and mechanisms of atmospheric escape are summarised in Section 1.4. This is logically followed by an exploration of enshrouded systems in Section 1.5. Star-planet interactions are the focus of Section 1.6. I provide an overview of the key achievements of this thesis in Section 1.7.

1.1 A Nobel pursuit

Less than thirty years ago it was unclear whether the Sun was unique in hosting planetary bodies. The detection of Jupiter-sized planet 51 Pegasi b orbiting a main sequence star (Mayor & Queloz, 1995) heralded a new era of astronomy. This discovery would be awarded the 2019 Nobel Prize in Physics.

Today, the search for exoplanets is arguably the most exciting and dynamic field in astronomy. In excess of 4200 exoplanets have been detected[1], more than half by the Kepler Space Telescope (Kepler: Borucki et al., 2010). Launch of the next generation of planet-hunting telescopes, such as the Transiting Exoplanet Survey Satellite (TESS: Ricker et al., 2014) in 2018 (which already has in excess of 2000 candidates) and PLAnetary Transits and Oscillations of stars (PLATO: Rauer et al., 2014) due in 2026, signals that the 'Exoplanet Goldrush' is under way.

Two particularly notable detections during the course of this research project involve red dwarf hosts: Proxima Centauri b is an Earth-like planet orbiting within the habitable zone of our nearest solar system (Anglada-Escudé et al., 2016). Seven Earth-like bodies orbit ultra-cool dwarf star TRAPPIST-1, three of which reside within the habitable zone (Gillon et al., 2017). These discoveries have cemented Exoplanetology centre stage in the astronomical landscape.

The discipline is becoming increasingly sophisticated. Recent work seeks to characterise discovered worlds, analysing the chemical make-up of their atmospheres and even bulk composition (e.g. Lopez-Morales, 2010; Parmentier et al., 2016; Sing et al., 2016; Bell & Cowan, 2018). Instruments aboard the CHaracterising ExOPlanet Satellite (CHEOPS: Benz et al., 2018), Atmospheric Remote-sensing Infrared Exoplanet Large-survey (ARIEL: Tinetti et al., 2016) and the James Webb Space Telescope (JWST: Gardner et al., 2006) will provide unprecedented insight into the chemistry and physical processes of alien worlds.

This work is undertaken as part of the 'Dispersed Matter Planet Project' (DMPP), the underlying hypothesis of which is that short period, close-in exoplanets are stripped of material which is subsequently deposited in dispersed circumstellar gas clouds.

[1] http://exoplanet.eu/: accessed 2 July 2020

By searching for absent flux in stellar spectra that has been absorbed by the ablated material, the DMPP has developed an efficient method for identifying short-orbit planet hosts (Haswell et al., 2019; Staab et al., 2019; Barnes et al., 2019).

This thesis considers the relationship between host stars and close-orbiting planets. In particular, I probe whether the atmospheres of close-in planets are ablated by intense irradiation from their hosts and, if so, what happens to lost material. I also explore how the magnetic and tidal interactions between bodies manifest as observable signals. By focusing on the most extreme exoplanetary systems, I have gained insight into physical processes and interactions that are likely ubiquitous in all star-planet systems but that are difficult to observe. I consider the final stages of planetary life and, in doing so, explore the ultimate fate of Planet Earth.

Figure 1.1 plots the celestial coordinates of confirmed exoplanet hosts. Kepler's field of view—apparent as a dense Maltese cross—illustrates what can be found when we look intensively at a given region of space. There remains much to discover...

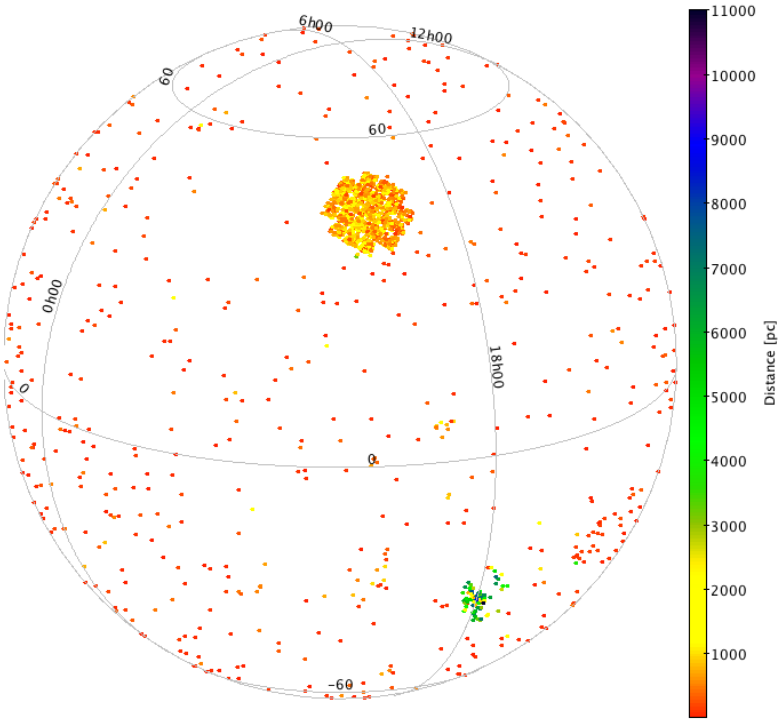

FIGURE 1.1: A sky plot of the celestial coordinates of confirmed exoplanet hosts from Exoplanet.eu, along with a colour scale showing distance to the host from Earth. Kepler's field of view is apparent as a dense, Maltese cross-like region. More distant systems that are detected by microlensing, cluster at the galactic bulge.

1.2 The known exoplanet population

1.2.1 Detection

A variety of techniques have been successfully deployed to detect exoplanets. A star undergoes a reflex motion about the star-planet barycentre during planetary orbit. This results in a radial velocity (RV)—being the component of a star's motion along the line of sight—that varies with time. The RV technique involves accurately measuring stellar radial velocities. It was used to discover the first exoplanet (Mayor & Queloz, 1995), and proved to be the most successful detection method during the late 1990's and 2000's.

The transit technique has subsequently become the most prolific detection method, primarily due to the success of the Kepler missions. A planet that travels between the disc of its host star and an observer attenuates light from the host, causing a periodic drop in apparent flux. This manifests as a characteristic dip in the star's light curve. The first exoplanet transit detections, observed by Henry et al. (2000) and Charbonneau et al. (2000), provided confirmation that Jupiter-mass planets in close-in orbits are comparable to the gas-giants of our solar system. When a planet transits, stellar light passes through the planetary atmosphere. Where the planet atmosphere strongly absorbs star light, transit depth increases. Some light is absorbed by atoms, molecules and condensates present in the atmosphere (Brown, 2001). Transmission spectroscopy involves obtaining a wavelength dependent spectrum of this light in order to analyse the chemical make-up of the planetary atmosphere.

The transit technique provides an estimate of planet radii. If mass has been deduced from the RV technique, then the planet's density and composition may be estimated. Planets detected by both transit and RV methods are particularly valuable as bulk density may be estimated, providing insight into planet composition.

Other successful planet-finding methods are astrometry, direct imaging, microlensing, and timing based techniques that involve pulsars and transit timing variations. The reader is referred to Perryman (2018) for a detailed description of planet-hunting methods.

As of 2 July 2020, 4281 exoplanet detections have been confirmed[2]. Of these, 2341 have been discovered by Kepler and 52 by TESS[3]. Table 1.1 ranks each detection technique by the number of confirmed discoveries.

Technique	Discoveries
Primary transit	3051
Radial velocity	888
Direct imaging	139
Microlensing	122
Pulsar	43
Transit time variation	19
Astrometry	11
Unspecified	8
Total	4281

TABLE 1.1: The number of exoplanet discoveries with corresponding discovery method at 2 July 2020 (Exoplanet.eu).

Planet distributions

Figure 1.2 shows the period-mass and period-radius distributions for confirmed exoplanets, along with the discovery method. Clustering in the plots results from selection bias: the transit method tends to find close-in, *giant* planets, while the RV technique preferentially discovers *massive* orbiting bodies. The planets with the longest orbital periods are found by direct imaging and microlensing. A feature known as the 'Neptune Desert' (described in Section 1.2.3) is evident in both distributions at short periods ($P < 3$ d). Three distinct groups have been identified in the period-mass-metallicity diagram: hot Jupiters, Jupiter-like planets at long periods, and the low mass planet population (Sousa et al., 2019).

[2] http://exoplanet.eu/: accessed 2 July 2020
[3] https://exoplanetarchive.ipac.caltech.edu/index.html: accessed 2 July 2020

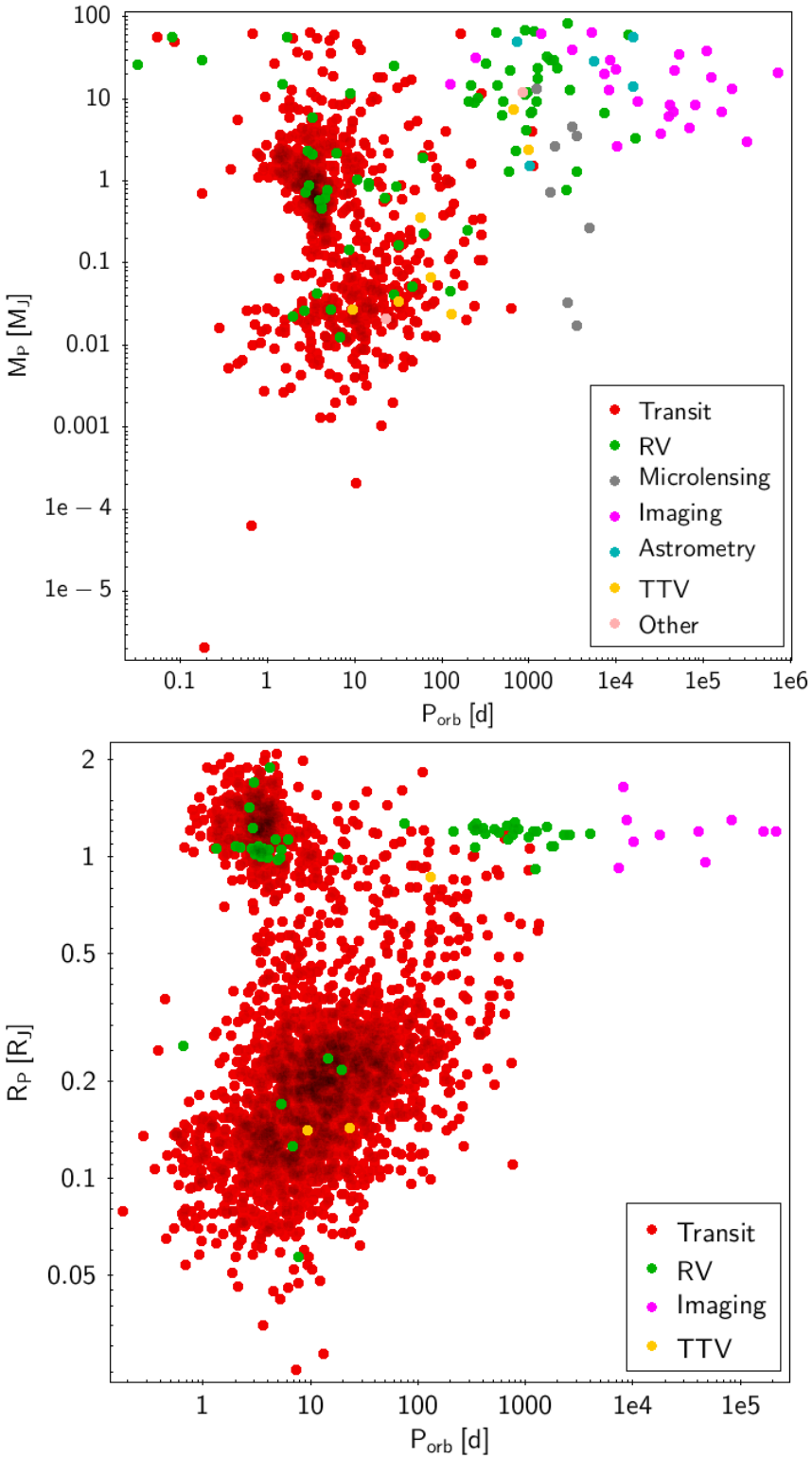

FIGURE 1.2: Top panel: period-mass distribution for confirmed exoplanets with marker colours indicating discovery method. Bottom panel: period-radius distribution. Clustering reflects selection bias. (Plotted using data from Exoplanet.eu.)

1.2.2 Close-orbiting systems

Due to selection bias of the most successful planet detection techniques, many exoplanets have been identified in short-period orbits ($P_{\rm orb} < 10$ d). In this section, I consider the classes of planet found in close-in orbits.

Hot Jupiters

Hot Jupiters (HJs) are gas giants of mass $0.36 - 11.8\ M_{\rm J}$ in $1.3 - 111$ d orbits around their hosts (Winn et al., 2010). They make up only $\sim 1\%$ of confirmed planets—a statistic likely inflated by detection bias as close-in, giant planets are the easiest to find with the transit and radial velocity techniques (Mordasini et al., 2009). In the Kepler population, 0.43 ± 0.05 % of stars are expected to host HJs with $P_{\rm orb} < 10$ d (Fressin et al., 2013). Occurrence rates around smaller stars are even lower: 0.07 % of M-dwarfs host planets with mass $1 - 10\ M_{\rm J}$ (Meyer et al., 2018). Several HJs have been found orbiting M dwarfs, including Kepler-45b (Johnson et al., 2012), HATS-6b (Hartman et al., 2015) and NGTS-1b (Bayliss et al., 2018).

HJs are believed to have formed far out from their hosts, before migrating inward and halting. Teyssandier et al. (2019) describes HJ formation through secular chaos and dynamical tides. Binarity appears to play a key role: in a sample of 38 systems with $M_{\rm P} > 7\ M_{\rm J}$ orbiting within 1 AU, $\sim 79\%$ have outer companions on separations between $20 - 10000$ AU, with a preference for wide binaries (Fontanive et al., 2019). There is also a higher binary rate for more massive and shorter period planets. However, not all HJs halt in time: Hamer & Schlaufman (2019) find that HJs hosts have a smaller Galactic velocity dispersion than non-HJ hosts, implying the HJ population is younger. Tidal interactions may cause inward spiral and ultimately destruction while on the main sequence.

Ultra-hot Jupiters

Ultra-hot Jupiters (UHJs) are gas giants with dayside temperatures exceeding 2200 K in short-orbital periods, typically around early-type stars (Parmentier et al., 2018). They are likely rare (Wright et al., 2012) but several have been observed, including KELT-9b (Yan & Henning, 2018), MASCARA-2b (Casasayas-Barris et al., 2018), WASP-12b (Fossati et al., 2010), WASP-18b (Arcangeli et al., 2019) and WASP-103b (Kreidberg et al., 2018). High temperatures mean atmospheric constituents are gaseous on their day-side, making UHJs suitable for transmission spectroscopy (Parmentier et al., 2016). They are ideal targets for thermal emission measurements as their peak wavelengths approach stellar values (Parmentier et al., 2018). UHJs are expected to be tidally locked. Incident flux is mainly re-emitted from the day-side rather than being reflected or circulating to the night-side (Torres et al., 2005; Deming et al., 2005; Désert et al., 2011). Temperature gradients of > 1000 K between the day- and night-side of UHJs are predicted (Perez-Becker & Chiang, 2013; Komacek & Showman, 2016; Schwartz et al., 2017; Zhang et al., 2018).

The day-side atmosphere of UHJs resemble those of stellar atmospheres due to the dissociation of molecular constituents. While most hot giants have atmospheres largely constituted of molecular hydrogen (H_2), at day-side UHJ temperatures H_2 will dissociate. Due to the contrast between day and night time temperatures, the local thermal equilibrium (LTE) H_2 dissociation fraction also varies (Bell & Cowan, 2018). Atomic hydrogen from the day-side will transfer energy to the night-side, whereupon it will recombine. Recombination of H into H_2 is an extremely exothermic process, which increases heat recirculation efficiency in the UHJ atmosphere. This may influence atmospheric escape from these planets. Seidel et al. (2019) identify a broadened neutral sodium feature in the atmosphere of WASP-76b, which they attribute to super-rotation in the upper atmosphere. Turner et al. (2020) detect ionized calcium (Ca II triplet) absorption in the atmosphere of KELT-9b, the hottest known planet with a day-side temperature of > 4500 K.

Super-puffs

Super-puffs are low density ($\rho \leq 10^{-1}$ g cm^{-3}) planets with $M_P \leq 5\,M_\oplus$ and $R_P > 5_\oplus$. Examples include Kepler 51c, Kepler 51d; Kepler 79d, Kepler 79e (Jontof-Hutter et al., 2014) and Kepler 87c (Ofir et al., 2014). Due to low surface gravity, super-puffs likely experience high mass-loss rates, even without strong stellar radiation, and should disperse over $\sim 10^3$ years (Wang & Dai, 2019). Super-puffs are ideal targets for transmission spectroscopy due to high scale heights.

Ultra-short period planets

Ultra-short period (USP) planets have radius $R_P < 2\,R_\oplus$ and sub-day orbits. Sanchis-Ojeda et al. (2014) identified ~ 100 such planets in Kepler transit data, finding planets with periods as short as 4 hours. The occurrence rate of USP planets within the known population of exoplanets is $\sim 1\%$, which is similar to HJ occurrence. While the one-day cut-off was arbitrarily chosen by Sanchis-Ojeda et al. (2014), evidence has emerged that sub-day orbital period planets have a distinct period distribution compared to $P_{\rm orb} > 1$ d planets (Petigura et al., 2017b; Lee & Chiang, 2017; Petigura et al., 2017b; Winn et al., 2018). There are 50% more planets just below the $P_{\rm orb} = 1$ d boundary than just above it.

USP planets have other distinct properties: small radii, with most $1R_\oplus \lesssim R \lesssim 1.4R_\oplus$ (Winn et al., 2018), likely resulting from photo-evaporation or "boil-off" due to exposure to extremely high irradiation levels (see Section 1.4.2). USPs have higher mutual inclinations than other Kepler planets (Dai et al., 2018). In contrast to HJs, they are frequently observed in multi-planet systems. However, for F, G and K hosts, USPs have eight times fewer co-transiting external companions compared to short-period (SP) planets (Petrovich et al., 2019). Period ratios between USP planets and their closest companion (P_1 / P_2) are an order of magnitude higher than period ratios in other Kepler multiple planets systems. Rappaport et al. (2013) found that most extreme USP planets have an iron-like density.

The formation and migration of USP planets remains poorly understood (Adams et al., 2016; Lopez & Rice, 2018; Pu & Lai, 2018). USP planets may either have formed *in situ* (Chiang & Laughlin, 2013) or migrated to their close-in position from more distant orbits (Ida & Lin, 2004; Schlaufman et al., 2010; Terquem, 2014) before loosing their H/He atmospheres to photo-evaporation (Valencia et al., 2010; Owen & Wu, 2013). Pu & Lai (2018) find that a low-eccentricity migration mechanism—where an inner planet maintains eccentricity due to secular forcing by external planets, then suffers tidal dissipation and orbital decay—more robustly produces the observed USP population than other proposed formation mechanisms. Using planet migration models, Carrera et al. (2019) find that small planets locked into mutual mean motion resonances migrate inwards in a chain. When the resonant chain reaches in the inner edge of the protoplanetary disk, its net inward torque forces the entire chain into closer orbits, whereas a single planet may been repelled by an outward force.

Disintegrating rocky planets

The catastrophically disintegrating rocky planet Kepler-1520b was reported by Rappaport et al. (2012). Time variable photometric signals indicated that sublimation of the surface formed of a metal rich vapour, which condensed into dust particles. Other disintegrating systems have been identified including K2-22 (Sanchis-Ojeda et al., 2015), WD 1145+017 (Vanderburg et al., 2015a) and KOI 2700 (Rappaport et al., 2014). Escaping dust particles can have wavelength-dependent transmission spectra due to Mie scattering (Sanchis-Ojeda et al., 2014; Bochinski et al., 2015). Ridden-Harper et al. (2019) search for absorption by gas lost from K2-22 b but were unable to detect Na or Ca^+. This indicates either gas loss is lower than expected or, as Na or Ca^+ are likely accelerated to high velocities (450 km s^{-1} and 135 km s^{-1} respectively) by stellar wind and radiation pressure, the features may be broad, blue-shifted and difficult to detect.

1.2.3 Demographic features

Several remarkable features have emerged in the short-period planet population.

The Neptune Desert

Observations indicate an absence of Neptune-like planets on short-period ($P_{\rm orb} \lesssim 3$ d) orbits (Beaugé & Nesvorný, 2013). The feature has been identified in both RV (Zucker & Mazeh, 2002; Gaudi, 2005; Cumming et al., 2008) and transit (Fressin et al., 2013; Howard et al., 2013) data. The region—commonly called the sub-Jovian desert (Szabó & Kiss, 2011) or the Neptune Desert (ND: Mazeh et al., 2016)—remains conspicuous in the current exoplanet population (Figure 1.2). HJs congregate at its upper boundary, while close-in super-Earths nestle along the low border. The region is mostly devoid of hot Neptunes. Mazeh et al. (2016) provide an analytic formulae for its borders, plotted in Figure 1.3. These boundaries depend on fundamental stellar parameters, such as effective temperature, metallicity, surface gravity and stellar mass, rather than on planetary parameters such as tidal forces, surface gravity or Roche-lobe filling factor (Szabó & Kálmán, 2019).

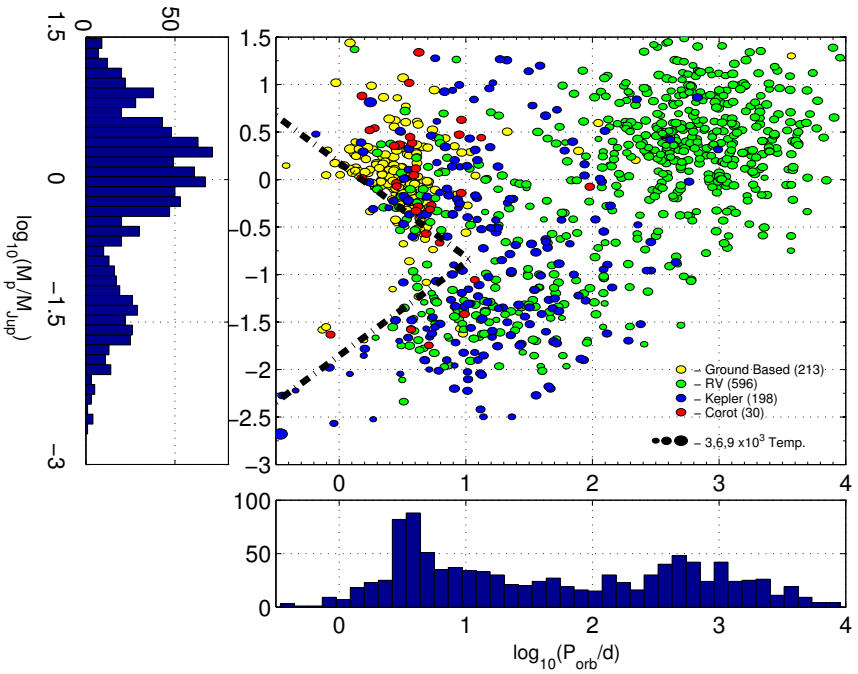

FIGURE 1.3: The period-mass exoplanet distribution with the boundaries of the Neptune Desert shown with dashed black lines. (Mazeh et al., 2016)

A variety of mechanisms to explain the ND have been suggested, which include: inward-migration of HJs that undergo rapid mass loss through photoevaporation and Roche-lobe overflow (Kurokawa & Nakamoto, 2014; Valsecchi et al., 2014); *in situ* formation of hot Jupiters by gas accretion onto super-Earth cores (Boley et al., 2016; Batygin et al., 2016); planet capture at or inside the disk inner edge (Beaugé & Nesvorný, 2013); planet arrival at the host star through planet-planet scattering, Kozai migration, or secular processes long after the protoplanetary disk has dispersed (Matsakos & Königl, 2016).

The Evaporation Valley

The 'evaporation valley' (EV) is a bimodal distribution of planet radii either side of a gap at $1.75 - 2\ R_\oplus$ (Fulton & Petigura, 2018; Van Eylen et al., 2018). The gap, which separates super-Earth from sub-Neptune planets, is thought to be caused by atmospheric escape. The feature was predicted by multiple models (e.g. Chen & Rogers, 2016; Owen & Wu, 2017; Lopez & Rice, 2018) and observed by Fulton & Petigura (2018) in a spectroscopic sample from the California Kepler Survey (Figure 1.4). Van Eylen et al. (2018) constrain the position of the EV, and characterise its slope with the power law $R \propto P^\gamma$ where $\gamma = -0.09^{+0.02}_{-0.04}$. The position of the EV is weakly dependent on the post-formation H/He envelope mass and evaporation strength (Mordasini, 2020). Low initial planet density—caused by high H/He mass—increases mass loss. Models that vary luminosity find similar valley locations (Owen & Wu, 2013; Jin et al., 2014; Lopez & Fortney, 2014; Chen & Rogers, 2016). The observed wide spread of $L_{\rm XUV}$ in young stars (Tu et al., 2015) does not shift the EV to the extent that it blurs away. Together, these factors make the EV a robust feature.

K2-36 is a K dwarf orbited by planets on either side of the EV. K2-36 b, with $R_{\rm P} = 1.43 \pm 0.08\ R_\oplus$ and $a = 0.022$ AU, has an Earth-like rocky composition and derived bulk density $\rho = 7.2^{+2.5}_{-2.1}$ g cm^{-3}. At a larger separation, K2-36 c, with $R_{\rm P} = 3.2 \pm 0.3\ R_\oplus$ and $a = 0.054$ AU, has a lower bulk density $\rho = 1.3^{+0.7}_{-0.5}$ g cm^{-3}, making it a sub-Neptune with a significant gas envelope. This system provides a fascinating test bed for the study of atmospheric escape.

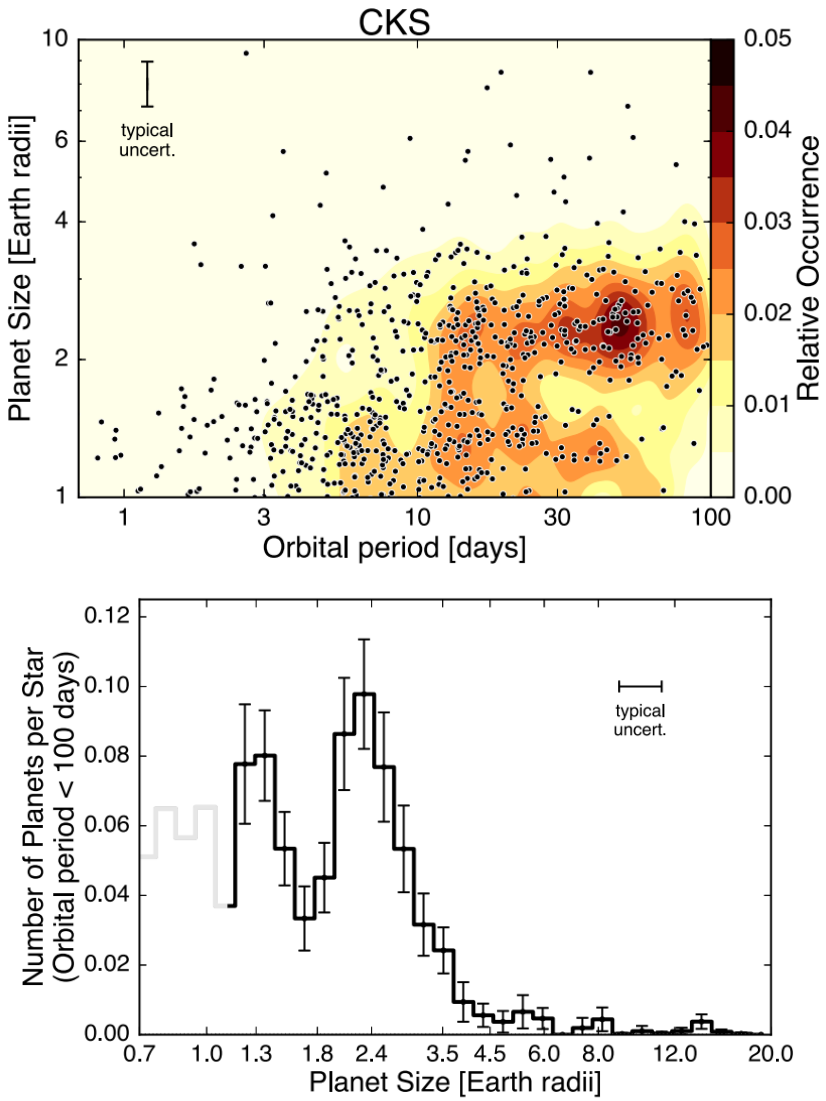

FIGURE 1.4: Top panel: a 2D planet radius–orbital period distribution. Bottom panel: planet radii where $P_{\rm orb} < 100$ d: the EV is apparent. (Fulton & Petigura, 2018)

The Mass-period Gap

Photoevaporation is expected to have a more modest effect on planet mass compared to radius. Nevertheless, Armstrong et al. (2019) report a gap in the mass-orbital period distribution for planets with $M < 20\ M_\oplus$ and $P_{\rm orb} < 20$ d. They suggest several mechanisms that could account for the gap, including: tidal star-planet interactions, disk-planet dynamical interactions, planet-planet interactions, or accreting material.

1.3 Stellar activity

1.3.1 Magnetic activity in stars

Stellar magnetic fields are generated by hydromagnetic dynamo processes in the overshoot zone, a region located between the convection and radiative zones (Steenbeck & Krause, 1966; Moffatt, 1978; Parker, 1979; Krause & Raedler, 1980; Miesch, 2012). The motion of plasma in the convective envelope across magnetic fields induces current, which in turn generates more field lines in a perpetual cycle. The magnetic fields are closed, buoyant, elastic, capable of being split, and can agglomerate. Under the influence of turbulent and helical motions in the convecting plasma, the magnetic field lines form twisted ropes or flux tubes (Steenbeck & Krause, 1966). Differential rotation—where the rotational velocity of the stellar medium is latitude dependent—stretches the ropes into poloidal and toroidal fields (Ossendrijver, 2003; Charbonneau, 2010). Dynamo-generated fields that rise to the stellar surface can become stressed by convective motions. These stresses are propagated into the chromosphere, where their release induces heating (Noyes et al., 1984).

1.3.2 Useful spectral lines

Several spectral lines are exceptionally useful for the study of stellar activity. The resonance lines of Ca II H & K, which occur at 3968.47 Å and 3933.66 Å respectively, are observed as strong absorption features in the atmospheres of cool stars and in the ISM, and as emissions from circumstellar shells. The widths of the H and K emission cores increase with rising luminosity (Wilson & Vainu Bappu, 1957). Evolved stars therefore have wider cores. The lines lie in the optical so are observable from the ground. Near ultra violet (NUV) wavelengths also contain strong resonance lines, including Mg II h & k lines at 2802.7 Å and 2795.5 Å respectively. As most NUV wavelengths are absorbed by the Earth's atmosphere, Mg II h & k lines are most commonly observed by space telescopes (e.g. the Hubble Space Telescope). The lines are established measures of chromospheric activity and are usually bright in active stars (Haswell et al., 2012).

Ca II H & K and Mg II h & k lines are used to measure stellar activity because they are abundant in stellar atmospheres and present strong resonance lines. The ionisation potentials of magnesium (7.6 eV) and calcium (6.1 eV) are significantly lower than hydrogen (13.6 eV), leading to both strong absorption and emission lines in the spectra of magnetically active stars.

1.3.3 Chromospheric activity

Figure 1.5 shows the contrasting emission profiles of an active and quiet Sun. Emission in the Ca II H & K lines is observed in stars that experience subsurface convection, which generally have of mass $M_\star < 1.5\ \mathrm{M}_\odot$ (Hoyle & Wilson, 1958; Schrijver et al., 1992). The intensity of Ca II H & K increases in response to the amount of non-thermal heating of the chromosphere (Noyes et al., 1984). There are two non-thermal mechanisms: magnetic and acoustic heating (Narain & Ulmschneider, 1996).

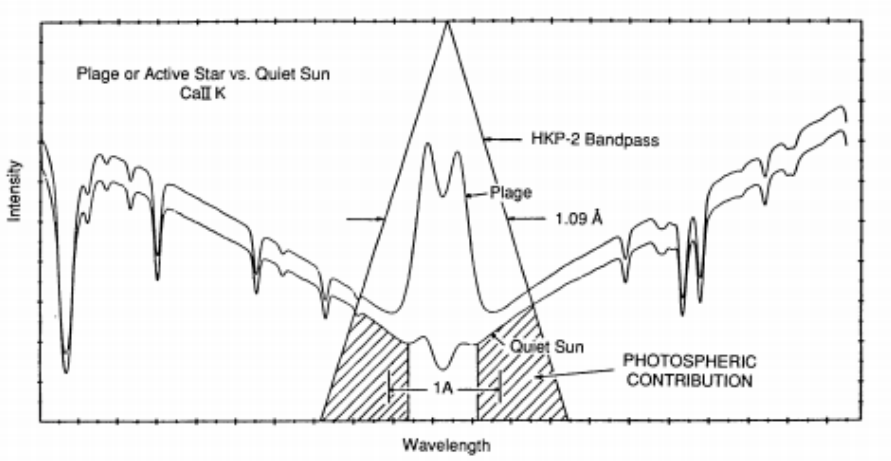

FIGURE 1.5: Bandpass of the HKP-2 spectrometer (being the second generation spectrometer put into operation at the Mount Wilson Observatory by Vaughan et al. (1978) specifically to measure Ca II H & K flux), with active and quiet Sun emission profiles (Duncan et al., 1991).

Magnetic heating of the chromosphere occurs when magnetic fields generated by dynamo processes rise to the stellar surface and are stressed by convective motions. These stresses propagate upward into the chromosphere, are released as heat, and result in temperature inversion and the formation of line core emission features (Noyes et al., 1984). The depth of the convection zone plays a role in magnetic field generation, while stellar mass and spectral type also dictate the properties of the stellar

convection zone. Ca II H & K flux is therefore a useful indicator of the strength of, and area covered by, magnetic fields (Leighton, 1959; Schrijver et al., 1989). The ubiquity of Ca II H & K emission (Eberhard & Schwarzschild, 1913) suggests that surface magnetic activity is a universal phenomenon in lower main sequence stars. Magnetic activity levels, and consequently chromospheric emission, may vary widely in convective stars (Staab et al., 2017).

Acoustic heating involves the dissipation of acoustic wave energy in the chromosphere, which generates basal flux emission (Buchholz et al., 1998).

The Mount Wilson system

The Mount Wilson project pioneered a system for measuring Ca II H & K emission line core strength (Wilson, 1968; Duncan et al., 1991). It established the 'S-index' as the standard metric of chromospheric activity, which has been measured for many stars (for example: Jenkins et al., 2011; Lovis et al., 2011a).

The S value is the ratio between Ca II H & K core emission flux and the flux in two specified continuum bandpasses (corrected for sky and instrument background):

$$S = \alpha \frac{H + K}{V + R}, \tag{1.1}$$

where H and K are the counts in the combined Ca II bands, and V and R are the counts in the violet and red continuum bands, which form in the photosphere (Middelkoop, 1982; Rutten et al., 1991; Barnes et al., 2016). The α term is a calibration factor that may be calculated daily or taken to be a constant value of 2.4 (Duncan et al., 1991; Baliunas et al., 1995).

The normalised flux in the H and K core bandpasses (R_{HK}) is:

$$R_{\mathrm{HK}} = \frac{F_{\mathrm{HK}}}{\sigma T_{\mathrm{eff}}^4}, \tag{1.2}$$

where F_{HK} is the total flux per cm^2 at the stellar surface in the H and K bandpasses, σ is the Stefan-Boltzmann constant, and T_{eff} is the effective temperature (Noyes et al.,

1984; Mittag et al., 2013). Figure 1.6 shows the $R_{\rm HK}$ distribution for a sample of unevolved field stars and planet hosts.

To correct for the fraction of total emission within H and K bandpasses that is photospheric rather than chromosheric in origin, Noyes et al. (1984) take the following step:

$$R'_{\rm HK} = R_{\rm HK} - R_{\rm phot}, \tag{1.3}$$

where $R_{\rm phot}$ is the photospheric contribution to the observed H and K line emissions. In order to minimise chromospheric contribution, $R_{\rm phot}$ is determined from the K line profiles of chromospherically weak lower main sequence stars. It is assumed that $R_{\rm phot}$ is independent of chromospheric emission, and depends only on $B-V$. The $R'_{\rm HK}$ term is the ratio of the chromospheric emission from the Ca II H & K cores to the total bolometric emission of the star. The parameter is approximately proportional to the fraction of non-radiative energy flux in the convective zone, which is converted to magnetic field-associated chromospheric heating. Converting S to $R'_{\rm HK}$ facilitates the comparison of the chromospheric activity of F, G and K stars.

Thus, three flux components contribute to measured S-values (Schrijver et al., 1989):

1. Photospheric flux (see the blue line Figure 1.6).
2. Non-variable basal chromospheric flux (the purple line Figure 1.6).
3. Time-variable, magnetic activity-related chromospheric flux.

As stars evolve, their photospheric contribution to the S-index changes, as does their magnetic activity levels. Activity metrics therefore show distinct distributions for different stellar classes and evolutionary phases (Mittag et al., 2013; Staab et al., 2017). While $\log(R'_{\rm HK})$ accounts for the colour-dependent photospheric contribution to core bandpasses, thus allowing F, G and K stars to be directly compared, it does not account for changes in magnetic activity as a star evolves. It is therefore necessary to separate main sequence, sub-giant and giant stars into distinct categories when analysing $\log(R'_{\rm HK})$ distributions.

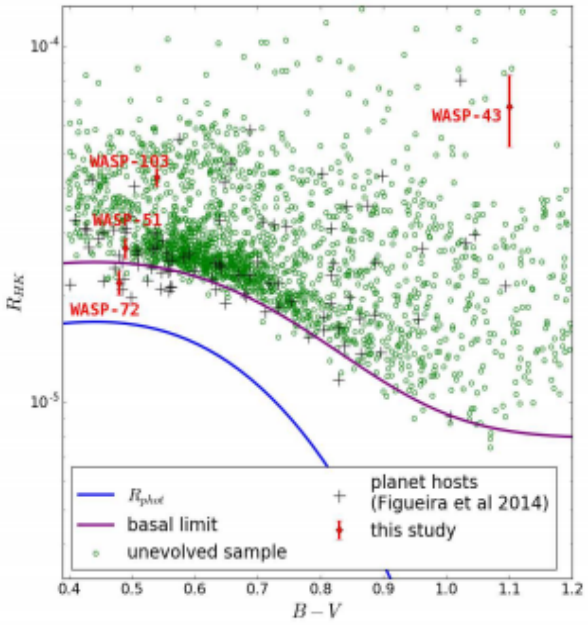

FIGURE 1.6: The R_{HK} distribution for main sequence planet hosts and field stars. The chromospheric basal limit and the photospheric contribution to R_{HK} (R_{phot}) are indicated. Staab et al. (2017)

Basal limit

The 'basal limit' is the minimum expected level of chromospheric activity in main sequence stars. During the Sun's activity minimum in 2008/9, when it was devoid of magnetically induced features such as plage and spots, Schröder et al. (2012) measured basal-level emission. The basal limit corresponds to this "quiet Sun"-like state, where stars are devoid of active regions. The basal limit is independent of the stellar dynamo and its activity cycles. Acoustics and small-scale magnetic mechanisms appear to stimulate basal activity (Hall, 2005; Buchholz et al., 1998). Placing a value on the basal limit is not straight forward. Wright (2004) found that targets with $\log(R'_{\mathrm{HK}}) < -5.1$ were either evolved stars or main sequence stars with unusually low activity. This value is adopted as the basal activity limit for main sequence F, G and K stars in this thesis, with recognition that it is a useful but arbitrary choice.

The Vaughan-Preston Gap

A bimodal $\log(R'_{\rm HK})$ distribution has been identified in large stellar populations, with active and inactive samples separated by the so-called 'Vaughan-Preston gap'—a region of depletion of stars of $B-V > 0.6$ and $-4.7 < \log(R'_{\rm HK}) < -4.9$ (Vaughan & Preston, 2002; Gray et al., 2006; Foukal, 2018). Various explanations have been invoked to explain the presence of the Vaughan-Preston (VP) gap, including: different modes of dynamo operation above and below the gap; a transition in dynamo behaviour at a critical Rossby number (Tobias, 1997; Brandenburg et al., 1998; Giampapa, 2015); or a spatial shift from large star spots to smaller facular and plage flux tubes as the magnetic dynamo weakens with age (Foukal, 2018). Boro Saikia et al. (2018) find that in a large sample of 4454 cool stars, the VP gap is less significant than previous studies suggest. They propose that this is due to the gradual spin down of main sequence stars from high to low activity with no sudden break of activity at intermediate levels.

1.3.4 Stellar rotation and chromospheric activity

Stars spin-down during their main sequence life time as angular momentum is lost via magnetized winds (Kawaler, 1988; Bouvier et al., 1997). Rotation period reflects angular momentum loss of the star. Gyrochronology considers the relationship between period, age and mass and is used to estimate the age of low-mass stars (Kawaler, 1988; Barnes, 2003, 2007; Mamajek & Hillenbrand, 2008).

Chromospheric activity is the product of turbulent convection and rotation. There is a strong correlation between stellar rotation rate and chromospheric activity levels in stars with subsurface convection zones (Kraft, 1967). As stars spin down with age, stellar chromospheric Ca II emissions decrease (Noyes et al., 1984; Wilson, 1963), meaning magnetic activity may be used as an age indicator (Skumanich, 1972; Soderblom et al., 1991; Reiners & Mohanty, 2012; Santos et al., 2016; Lorenzo-Oliveira et al., 2016). A smooth decrease in chromospheric activity is expected for decreasing age/rotation period (Soderblom et al., 1991; Mamajek & Hillenbrand, 2008; Lorenzo-Oliveira et al., 2016), as shown in Figure 1.7. Pace & Pasquini (2004)

and Pace (2013) suggest this age-activity relation holds only for the first 2 Gyr of a field star's life. In contrast, Lorenzo-Oliveira et al. (2018) find the age-activity relation is statistically significant for $6-7$ Gyr for solar-like stars.

The rotation history of host stars may be inferred from observed atmospheric properties of exoplanets. Sub-Neptune atmospheres in particular chart the XUV evolution of the host star. K stars HD 3167 and K2-32 have very different rotation histories, despite having very similar masses (Kubyshkina et al., 2019).

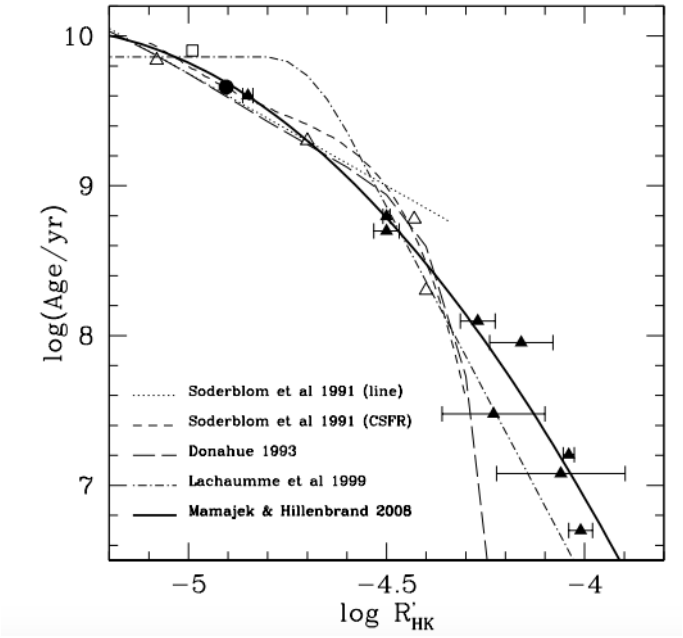

FIGURE 1.7: Mean $\log(R'_{\rm HK})$ cluster values (interpolated to solar $B-V$) plotted against cluster age from several studies detailed, with a best-fit polynomial as a dark solid line. The filled circle is the Sun. (Mamajek & Hillenbrand, 2008)

Rotational modulation and period

Magnetic activity coupled with differential stellar rotation may produce observable features, including starspots, faculae, spicules, filaments and plage. These features manifest as photometric and radial velocity variability, which is a source of nuisance known as "jitter" in planet searches. Hillenbrand et al. (2014) estimated that in a young active stellar population $\log(R'_{\rm HK}) = -4.0$ produces jitter levels $> 200\,{\rm m\ s^{-1}}$. Accounting for jitter is a key challenge in the hunt for terrestrial planets in the habitable zone. However, the features can also prove useful: starspots may last long

enough to be tracked during a star's rotation, so may be used to trace rotation. Stellar rotation periods have been estimated for a number of Kepler stars from rotational modulation of surface star spots (Walkowicz & Hawley, 2009; McQuillan et al., 2013a,b, 2014; García et al., 2014; Buzasi et al., 2016). McQuillan et al. (2014) use an autocorrelation function to detect rotation periods ranging from $\sim 0.2 - 70$ d in 34030 Kepler MS stars with $T_{\rm eff} < 6500$ K. They identify a bimodal rotation period distribution with peaks at 19 and 33 d, suggesting separate rounds of star formation.

1.4 Atmospheric escape

The possibility of atmospheric escape has been apparent since the discovery of highly irradiated gas giants (Mayor & Queloz, 1995; Burrows & Lunine, 1995).

1.4.1 Observations of atmospheric escape

Atmospheric escape may be observed during transit, with various probes available.

Lyman α

Neutral hydrogen can be detected through the Lyman-α transit, i.e. where a transit event is observed in the Lyman-α wavelength (1215.67Å). This has a large cross section at line centre (Hansen & Peng Oh, 2006). Lyman-α transits have been observed for a number of HJs: HD 209458b with a $\sim$ 15% flux dip (Charbonneau et al., 2000; Vidal-Madjar et al., 2003); HD 189733b with a $\sim$ 5% flux dip (Lecavelier des Etangs et al., 2010); and GJ 436b with a $\sim$ 56% drop in Lyman-α flux (compared to 0.69% in the optical: Figure 1.8), providing clear evidence of planetary mass-loss (Kulow et al., 2014; Ehrenreich et al., 2015; Lavie et al., 2017). An extended atmosphere of neutral hydrogen has also been observed around warm Neptune GJ 3470b, with H atoms escaping at a rate of 10^{10} g s^{-1} (Bourrier et al., 2018). This escape rate indicates the planet has lost $4 - 35$ % of its current mass over $\sim$ 2 Gyr. The atoms may be photoionized and blasted away from the star by radiation pressure, leading to a smaller exosphere than GJ 436b despite faster escape.

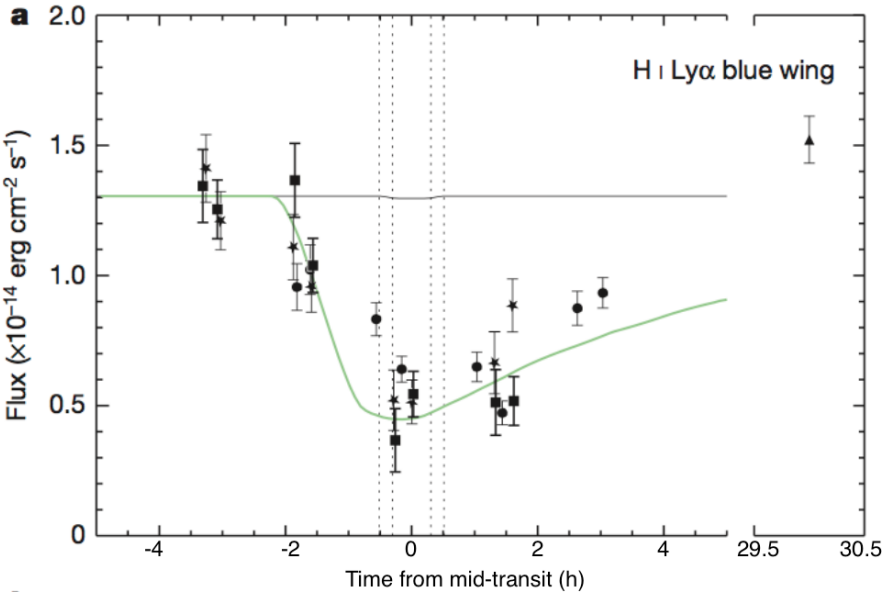

FIGURE 1.8: Optical transit of GJ 436b with 0.9% depth shown black. Lyman α with a mid-transit depth of $56.2 \pm 3.6\%$ in green. (Ehrenreich et al., 2015)

Helium triplet

In contrast to Lyman-α, the metastable helium triplet at 10833 Å is located in a spectral region devoid of strong ISM absorption with a bright stellar continuum (Seager & Sasselov, 2000; Indriolo et al., 2009; Oklopčić & Hirata, 2018). Absorption of 10833 Å photons excite He to the 2^3S level, yielding a significant transition signal (Oklopčić & Hirata, 2018). K-type stars hosting close-orbiting planets ($a \lesssim 0.05$ AU) provide the ideal conditions for exciting helium atoms (Oklopčić, 2019).

In a stunning observation, Spake et al. (2018) detected the He I triplet in the infrared transmission spectrum of low-density, warm Saturn WASP-107 b. They suggest an extended atmosphere eroding at a rate of $10^{10} - 3 \times 10^{11}$ g s^{-1}. A follow-up study confirmed detection of the feature and resolved it at high confidence (Allart et al., 2019). The He I triplet has also been detected in the spectrum of warm Neptune HAT-P-11 using high resolution, ground-based observations and the Hubble Space Telescope (HST) (Allart et al., 2018; Mansfield et al., 2018). A He loss rate of $\lesssim 2 \times 10^5$ g s^{-1} is estimated. The planet's upper atmosphere extends beyond 5 R_P. He I absorption is also reported in the transmission spectra of HJs WASP-69 (Nortmann et al., 2018) and HD 189733b (Salz et al., 2018).

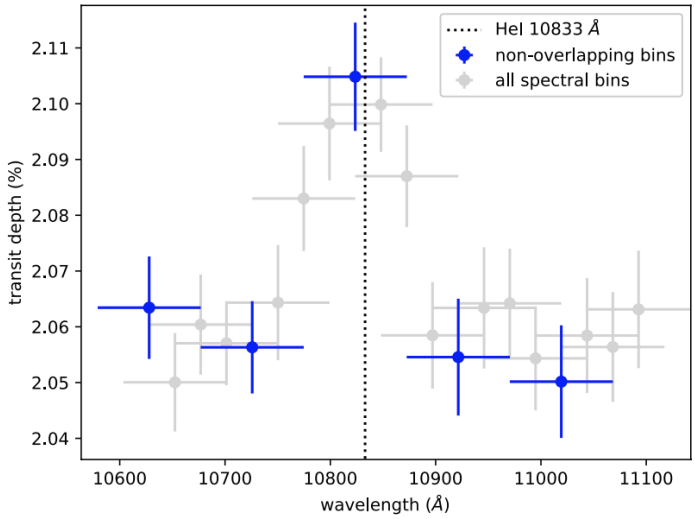

FIGURE 1.9: WASP-107b narrow-band transmission spectrum with a peak at 10833 Å (Spake et al., 2018).

UV and X-ray observations

Atmospheric escape may also be observed in the UV, where heavy elements in enlarged atmospheres indicate hydrodynamic escape in action (Section 1.4.2). Large transit depths in O I and C II lines of HD 209458 (Vidal-Madjar et al., 2003) and HD 189733 (Ben-Jaffel & Ballester, 2013) confirm hydrodynamic escape. Near-UV (NUV) transits of WASP-12b are three times deeper than in the optical, revealing extensive gas beyond the Roche lobe (Figure 1.10: Haswell et al., 2012). FUV observations can place constraints on mass-loss (Fossati et al., 2015a, 2018b). Cubillos et al. (2020) find evidence of ionized iron beyond the Roche lobe of inflated transiting HJ HD 209458b, indicating hydrodynamic escape can power heavy atoms beyond the Roche lobe even in less-irradiated planets. Neutral Fe I has recently been detected in the atmosphere UHJ WASP-121 (Gibson et al., 2020). Heavy elements absorb X-rays, so X-ray observations may also be used to observe mass loss. For example, Poppenhaeger et al. (2013) observed a 8 % X-ray transit of HD 189733b.

1.4.2 Mass loss mechanisms

There are a number of mechanisms by which atmospheres may be lost, which may act simultaneously.

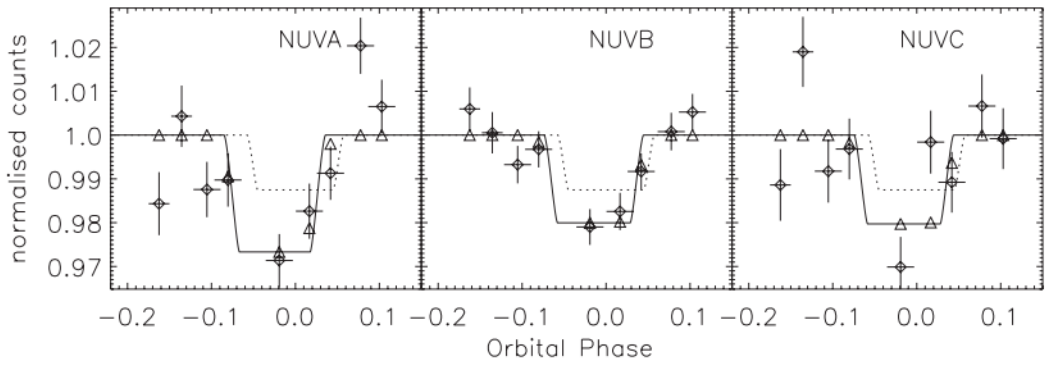

FIGURE 1.10: Best-fitting simple transit models (solid lines) for WASP-12's NUV light curves compared to optical transits (dotted line). (Haswell et al., 2012)

Non-thermal escape

Accreting planetesimal impacts can erode an existing atmosphere, with late stage residual planetesimal impacts having the potential to result in significant or catastrophic ablation. The impactor mass and atmospheric scale height determine the efficiency of atmospheric ejection (O'Keefe & Ahrens, 1985; Schlichting et al., 2015). Planetesimal impacts during the formation of hot Jupiters have a significant effect on atmospheric composition (Mordasini et al., 2016; Pinhas et al., 2016). Impact erosion could dominate mass loss on young terrestrial planets, with Jeans escape, thermal wind, and ablation by stellar winds unlikely to contribute significantly (Howe et al., 2020).

Photolytic dissociation may also act to unbind atmospheres. Photodissociation occurs when molecules are broken down by UV radiation and it can accelerate the loss of volatiles to space. Léger et al. (2004) find that atmospheric NH_3 in a planet 1 AU from a G2V star could be converted into N_2 and H_2 in 2 Gyr.

Thermal escape

For planets in close-in orbits, thermal processes are likely to dominate due to extreme heating. The Jeans escape parameter, being the ratio of gravitational to thermal energy in a gas, is key to determining the type of escape in action:

$$\lambda_{\rm J} = \frac{GM_{\rm P}\mu m_{\rm H}}{k_{\rm b}TR_{\rm P}}, \tag{1.4}$$

where G is the gravitational constant, μ is the mean particle mass in units of the mass of the hydrogen ($m_{\rm H}$), $k_{\rm b}$ is the Boltzmann constant, and T is the temperature of the upper atmosphere (Jeans, 1925; Chamberlain, 1963; Öpik, 1963). There are three main mechanisms of thermal escape:

Jeans escape

High energy molecules are liberated from planetary gravity when they exceed the escape velocity. This is known as Jeans escape, and is shown in context of the thermal Maxwell-Boltzman distribution in Figure 1.11. This mechanism defines the exosphere, which is the transition region between the atmosphere and interplanetary space. At the lower boundary of the exosphere—known as the exobase or thermopause—the molecular mean free path is equal to one pressure scale height, i.e. upward travelling particles have ballistic trajectories (Perryman, 2018). The Jeans parameter may be expressed as $\lambda_{\rm J} \equiv v_{\rm esc}^2/v_{\rm P}^2$, where $v_{\rm P}$ is the most probable velocity of the distribution and $v_{\rm esc}$ is the escape velocity at the exobase. Gas particles that exceed the escape velocity at the exobase are likely to escape. Particles with velocity $< v_{\rm esc}$ cannot escape. Early studies found hot Jupiters to be stable to classical Jeans escape, estimating a mass-loss of $10^{-4}\ M_{\rm J}$ per Gyr for 51 Peg (Guillot et al., 1996).

Boil-off

Fossati et al. (2017a) generalise the Jeans escape parameter for close-in, low density planets taking into account hydrodynamic and Roche lobe effects. They define a 'restricted' Jeans parameter Λ, and identify a 'boil-off' regime where escape is driven both by thermal energy and low planetary gravity. Planets with $\Lambda < 15 - 35$ experience boil-off. Atmospheres of hot ($T_{\rm eq} > 1,000K$) low-mass ($M_{\rm P} < 5\ M_{\oplus}$) planets with $\Lambda < 15 - 35$, shrink to smaller radii and out of the boil-off region.

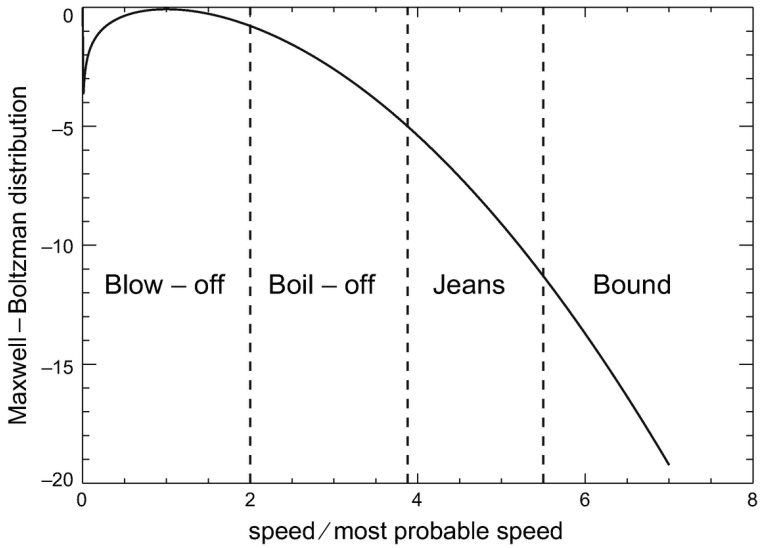

FIGURE 1.11: Atmosphere escape regimes discussed in the text along the tail of the Maxwell-Boltzmann distribution. (Linsky, 2019)

Hydrodynamic blow-off

Hydrodynamic escape is a thermally-driven form of Jeans escape where light atoms frictionally drag heavier species to super-escape velocities (Shizgal & Arkos, 1996). Continuous energy input is required at high altitude to maintain a high-velocity Maxwell-Boltzmann tail. Where the gravitational binding energy of the gas is surpassed by its internal energy ($\lambda_J < 1.5$), unimpeded atmospheric 'blow-off' results (Öpik, 1963). Significant hydrodynamic escape is predicted in systems where semi-major axis $a = 0.01 - 0.1$ AU. Atmospheres may nevertheless remain substantially unchanged for several Gyr (Guillot et al., 1996; Yelle, 2004; Tian et al., 2005).

Transition from Jeans $\rightarrow$ hydrodynamic escape

The relative altitudes of the scale height and the exobase dictate the dominant mass-loss mechanism. In a weakly irradiated atmosphere, the exobase lies below the scale height, with Jeans escape dominant and low mass-loss rates. For highly irradiated exoplanets, the atmospheric scale height $\rightarrow \infty$, with gas density and pressure becoming constant at large radii. The transition from Jeans to hydrodynamic escape occurs when the exobase is at the 'sonic point'—being a critical point where gas velocity of the hydrodynamic outflow reaches the speed of sound (Parker, 1956; Owen & Jackson, 2012). Flow beyond the sonic point is super-sonic, meaning information cannot

be communicated back below the exobase. So even if the outflow interacts with magnetic fields or the stellar wind, flow from the planet will remain unaffected. For atomic gas, the transition from Jeans to hydrodynamic escape occurs at $\lambda_J \sim 2 - 3$.

1.4.3 Evolution of mass-losing planets

Understanding mass-loss timescales as a function of mass explains planetary evolution, as well as the presence of the ND and EV (Owen, 2019).

Hot giant planets

Early studies of HJ mass-loss derived high hydrodynamic escape rates of up to 10^{12} g s^{-1} (Lammer et al., 2003). Hébrard et al. (2003) suggest that the H/He envelope of CoRoT-7b may have been stripped, leaving a terrestrial planet-like rocky or metallic core. Inward migration and Roche lobe overflow may compound escape, leading to atmospheric inflation (Trilling et al., 1998, 2002; Baraffe et al., 2004; Jaritz et al., 2005; Erkaev et al., 2007; Hubbard et al., 2007). Such processes may have sculpted hot Neptunes GJ 436 b and 55 Cnc e (Baraffe et al., 2005a,b). Roche-lobe overflow and rapid mass loss from HJs may result in a dichotomous population (Kurokawa & Nakamoto, 2014). Other simulations suggest lower escape rates of $\sim 10^{10} - 10^{11}$ g s^{-1} (Murray-Clay et al., 2009; Yelle, 2004; Murray-Clay et al., 2009; Owen & Jackson, 2012) and that HJs may lose only a small percentage of their fledgling atmospheres (Hubbard et al., 2007; Owen & Wu, 2013; Jin et al., 2014). Winn et al. (2017) argue that these findings indicate that close-in, low mass planets do not evolve from HJs. However, the latest models suggest that mass loss can be significant in *certain circumstances*. For example, Locci et al. (2019) find that low mass HJs with $P_{\rm orb} < 2.5$ d experience significant mass loss during their first 1 billion years. Allan & Vidotto (2019) find a 1 M_J planet would lose 1% its initial mass over 5 Gyr compared to 20% for a 0.5 M_J planet.

Ultra hot Jupiters

KELT-9b is in a 1.5 d orbit around a very hot A0-type star and has day-side temperature of > 4500 K. It could be stripped of its envelope during its main sequence lifetime (Gaudi et al., 2017). Fossati et al. (2018b) estimate a mass-loss rate of $10^{10} - 10^{11}$ g s^{-1}. Deeper than expected absorption lines for Mg I, Fe I, Fe II, T I, Na I, Cr II, Sc II and Y II have been detected in the transmission spectrum of KELT-9b, suggesting out flowing material is transported to higher altitudes (Hoeijmakers et al., 2019). Turner et al. (2020) detect ionized calcium (Ca II triplet) and a prominent $H\alpha$ absorption in the atmosphere of KELT-9b, confirming enlargement of the hydrogen envelope (Yan & Henning, 2018; Cauley et al., 2019b). WASP-33, which orbits a cooler star, experiences high XUV fluxes that power mass loss of 10^{11} g s^{-1} (Fossati et al., 2018b). Vines et al. (2019) estimate a mass loss rate of 10^{11} g s^{-1} from NGTS-6b—a $1.33 M_{\rm J}$ planet in a 21.17 h orbit around a metal-rich star. Even higher mass loss rates are expected from less massive planets orbiting EUV-bright intermediate-mass stars (Fossati et al., 2018b). Giant exoplanets around hot stars may experience catastrophic mass loss via Balmer-driven escape if $M_{\rm P} < 1 - 2\ M_{\rm J}$, and they orbit closer-in than KELT-9b (García Muñoz & Schneider, 2019). These findings challenge the assumption that giant planets are stable to atmospheric escape.

Low mass planets

Due to their large radii at young ages (Lopez et al., 2012; Lopez & Fortney, 2013), young low-mass planets receive high energy fluxes and experience the majority of their mass loss during the first 100 Myr of life (Owen & Wu, 2013; Lopez & Fortney, 2013). Thereafter, stellar flux decays and the planet contracts. If the period over which peak mass loss occurs is less than 100 Myr, the planet's envelope is stripped and only a bare core remains—as is seen for short-period Kepler planets (Dressing et al., 2015). If the peak mass-loss period is over 100 Myr, then the planet will retain a proportion of its atmosphere after its initial exposure to high energy fluxes. This leads to a bifurcation in the population—the 'Evaporation Valley' (Figure 1.12).

Core-powered mass loss models—where only cooling luminosity from rocky cores drives atmospheric escape—can recreate the position, shape, slope and peak magnitudes of the EV (Ginzburg et al., 2016, 2018; Gupta & Schlichting, 2019). These features are independent of stellar mass, age and metallicity (Gupta & Schlichting, 2020). Loyd et al. (2020) find no evidence that favours photoevaporation or core powered mass loss as the main cause of the EV. An integrated model is clearly required to probe this dynamic.

Planets in the galactic bulge may lose several Earth masses of atmosphere due to radiation from the Milky Way's central super massive black hole, Sagittarius A* (Wisłocka et al., 2019). Planets at a safe distance (> 7 kpc) do not undergo erosion. In tight planet-hosting, binary star systems tidal interactions between the stars can increase the amount of XUV energy output over the stellar lifetimes by up to $\sim$ 50 times, especially where separation $\lesssim$ 0.12 AU. This can significantly increase atmospheric losses from planets orbiting the stars (Johnstone et al., 2019).

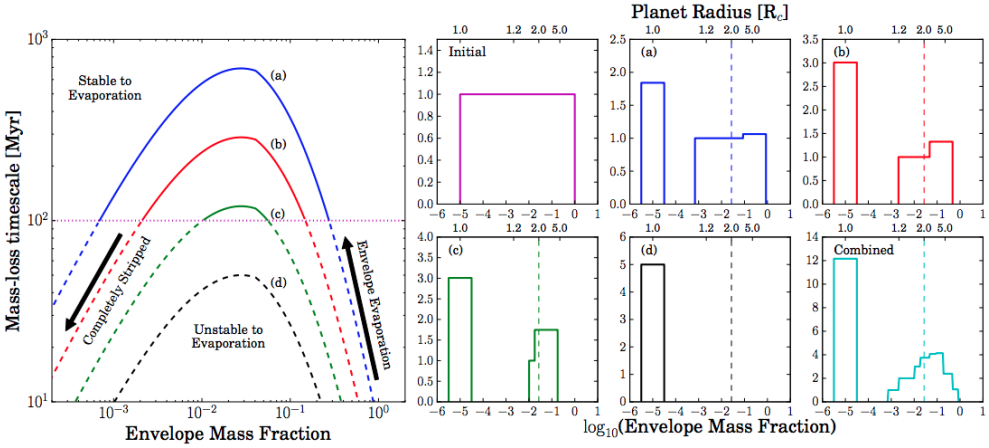

FIGURE 1.12: Schematic showing the mass-loss period and envelope mass fraction (left). Separation decreases for Models (a) to (d). Where mass-loss period $<$ 100 Myr (dashed lines), atmospheres are unstable to evaporation. Histograms (right) show the resulting envelope mass fractions and planet distributions. (Owen & Wu, 2017)

1.5 Enshrouded systems

Mass loss is likely ubiquitous in close-orbiting systems. Material stripped from the planet could be accreted onto the host star, lost to interplanetary space, or remain *in situ* within the star-planet system. The study of UHJ WASP-12b sheds light on the fate of stripped planetary material.

1.5.1 WASP-12b

WASP-12 is an F-type, main sequence star of age < 2.65 Gyr (Fossati et al., 2010). It hosts WASP-12b, a highly irradiated UHJ that orbits approximately one stellar diameter from its companion (Hebb et al., 2009). An enhanced near-UV (NUV) transit, caused by absorption of metal resonance lines in WASP-12b's exosphere, has been observed using transmission spectroscopy (Figure 1.10). In addition, a permanent anomaly in WASP-12's NUV spectrum has been identified: a dearth of the normally bright emission cores in the Mg II h & k resonance lines (Haswell et al., 2012).

Figure 1.13 compares WASP-12's NUV spectrum with HD 102634 and HD 107213—stars of analogous age and temperature that exhibit typical emission cores in the Mg II resonance lines (Valenti & Fischer, 2005). The spectrum of Procyon—a younger, hotter, slowly rotating ($v \sin i = 3.5$ km s^{-1}) mid-F subgiant that is often employed as an inactive comparator (Schröder et al., 2009)—is also plotted. The Mg II line profiles match well, except that WASP-12 completely lacks emission line cores. Similarly, Fossati et al. (2013) found the Ca II H & K lines cores of WASP-12 significantly deeper than those of stars with otherwise similar properties (Figure 1.14).

These findings are unusual given the WASP-12's estimated age and spectral type (Guinan & Engle, 2008). Mg II emission cores are expected even for slowly rotating sub-giants (Ayres, 2010). Three scenarios were proposed to explain the WASP-12 anomaly:

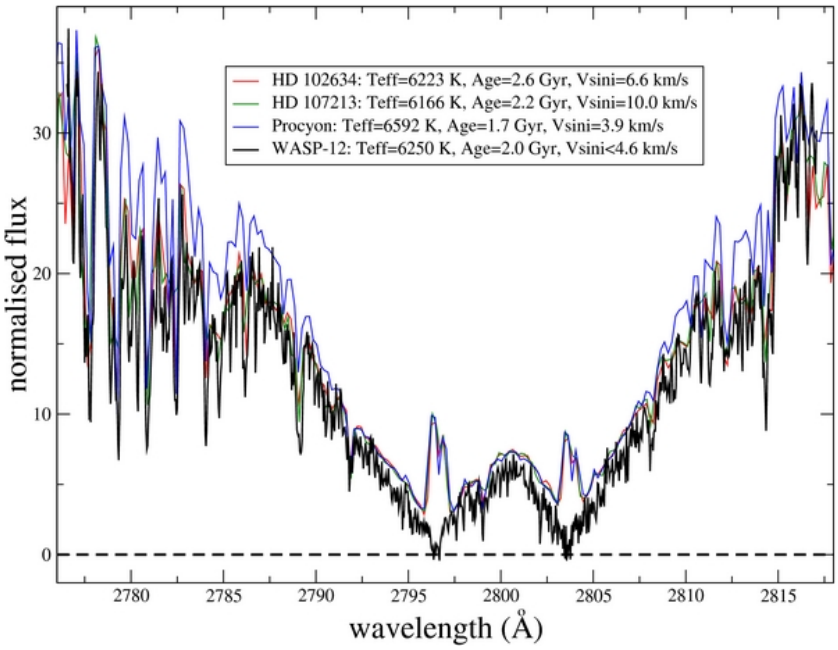

FIGURE 1.13: The Mg II h & k cores of WASP-12, HD 102634, HD 107213 and Procyon. No emission features are apparent in WASP-12's line cores. (Haswell et al., 2012)

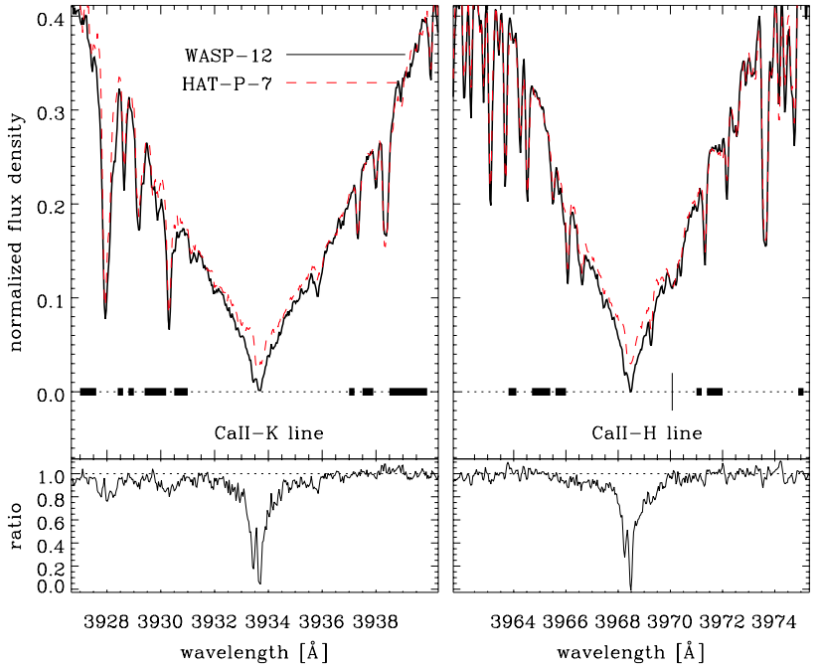

FIGURE 1.14: The Ca II H & K line profiles of WASP-12 and HAT-P-7. WASP-12 displays enhanced absorption in the inner line profile. (Fossati et al., 2013)

Low chromospheric activity

The absence of Mg II and Ca II in hot Jupiter host spectra may be attributed to low rotation velocities and a lack of stellar activity. This is, however, unprecedented for a solar-like stars such as WASP-12 (Fossati et al., 2013). Mg II and Ca II emission is expected even for slowly rotating sub-giants (Ayres, 2010; Haswell et al., 2012). HJ planets may suppress stellar activity through extreme tidal interactions (WASP-18: Miller et al., 2012; Pillitteri et al., 2014) but this seems unlikely here given WASP-12's mass. Figure 1.15 shows a plot of $B - V$ against $\log(R'_{\rm HK})$ for 50 transiting planet host stars (Knutson et al., 2010; Fossati et al., 2013). For the majority of stars $\log(R'_{\rm HK})$ values are between -4.6 and -5.0. WASP-12b however is seen as an extreme outlier, falling well below the basal level of chromospheric emission for late-type main sequence stars: $\log(R'_{\rm HK}) = -5.1$ for solar metallicity dwarfs (Henry et al., 1996; Wright, 2004). The spectra of normal stars that fall below the basal level must experience *external absorption* (Fossati et al., 2013).

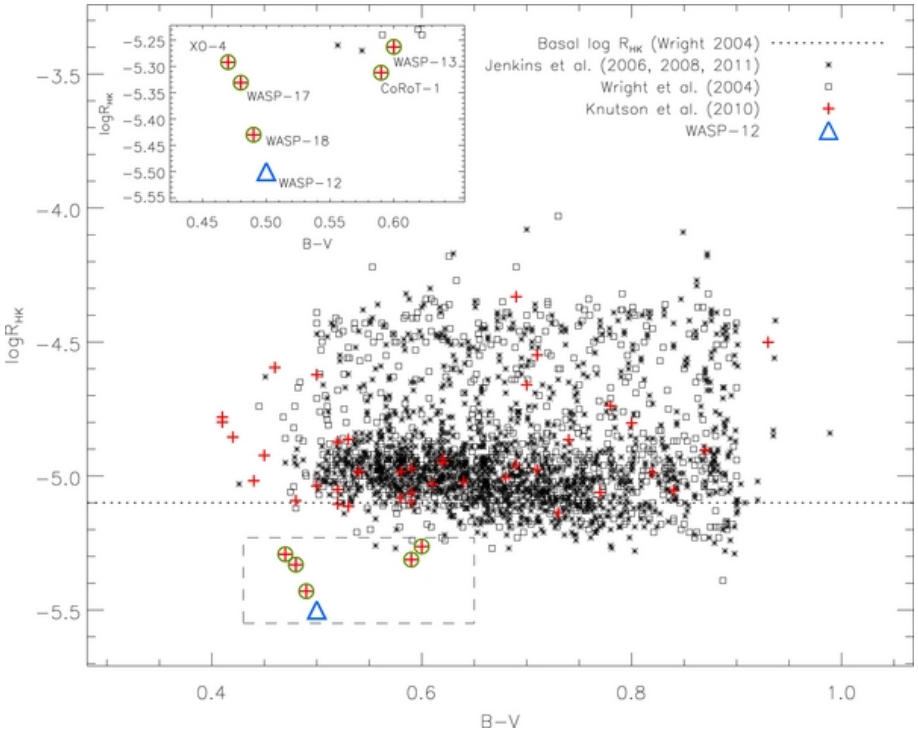

FIGURE 1.15: Colour-activity distribution for Wright (2004) sample and planet hosts. WASP-12 is the blue triangle, while circles indicate X0-4, CoRoT-1, WASP-17 and WASP-18. The dotted line is the basal limit. (Fossati et al., 2013)

Absorption by the interstellar medium

Absorption by the interstellar medium (ISM) can cause anomalous emission features. Measured activity may be depressed well below $\log(R'_{HK}) = -5.1$ for stars located more than 100pc from Earth (Fossati et al., 2015a). Haswell et al. (2012) calculated that the column density along the line-of-sight to WASP-12 would need to be 10 times higher than expected in order to fully absorb Mg II core emissions. Similarly Fossati et al. (2013) found that Mg and Ca ISM column densities in the direction of WASP-12 are far below the density required to produce full Mg II and Ca II line core depressions.

Absorption by ablated material in the WASP-12 system

The interpretation preferred by Haswell et al. (2012) is that gas stripped from the heavily irradiated UHJ WASP12b forms a spatially dispersed disk that enshrouds the system and possesses sufficient column density to absorb expected emission flux. The interpretation seems plausible and indeed has subsequently been supported by 3D hydrodynamic simulations (Figure 1.16: Matsakos et al., 2015; Carroll-Nellenback et al., 2017; Debrecht et al., 2018; Daley-Yates & Stevens, 2018). Harsh radiation may result in blow-off of the planet's atmosphere (Vidal-Madjar et al., 2004; Ehrenreich & Désert, 2011; Guo, 2011). Dwivedi et al. (2019) find that material from WASP-12b's upper atmosphere forms two streams that propagate towards and away from WASP-12, with mass loss rates reaching 10^{12} g s^{-1}. Outflows may be channelled by the planetary magnetic field and Roche equipotentials, while retaining the planet's angular momentum (Erkaev et al., 2007; Adams, 2011). The stripped material has an enhanced cross-sectional area to radiation, so may be forced outward from the planetary orbit to enshroud the system (Haswell et al., 2012). Gas may be expelled perpendicularly to the planet's orbit to distances more than twice the planet's diameter (Bisikalo et al., 2013). A relatively high temperature would prevent collapse of the disk back into the orbital plane (Fossati et al., 2013). The disk would be maintained by a continuous flow of fresh material from the close-orbiting planet.

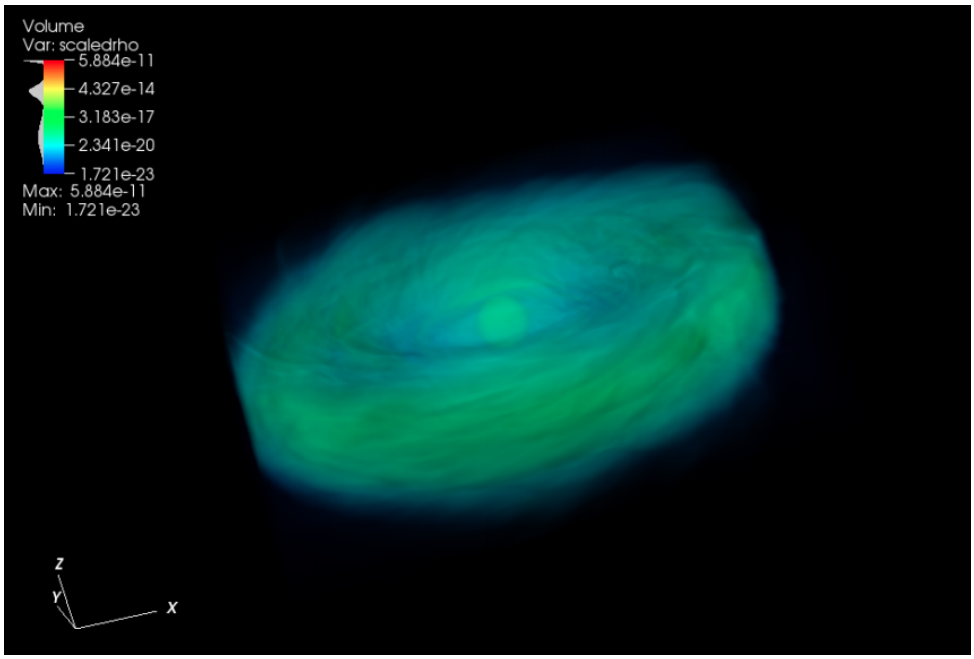

FIGURE 1.16: A 3D simulation showing the formation of a torus-like diffuse cloud around WASP-12. The colour code gives the hydrogen density in g cm^{-3}. The disk enshrouds the stellar surface as seen by an observer lying in the orbital plane. (Debrecht et al., 2018)

1.5.2 The Dispersed Matter Planet Project

Highly irradiated close-orbiting exoplanets may show anomalously low activity due to absorption by circumstellar gas clouds replenished by atmospheric escape from abating planets. This insight has provided the Dispersed Matter Planet Project (DMPP) with an efficient technique for identifying nearby compact planetary systems (Haswell et al., 2019). Upon identifying around 40 systems from $\sim$ 3000 nearby, bright, main sequence stars with lower-than-expected activity, the group have discovered close-in planets wherever $>$ 60 high precision RV measurements have been made.

DMPP-1 is a multi-planet system containing three super-Earths in $<$ 10 d orbits, and a warm Neptune in a 20 d orbit around F8V star HD 38677 (Staab et al., 2019). DMPP-2 is a $M_{\rm P} \sin i = 0.469\ M_{\rm J}$ planet in a 5.207 d orbit around the γ-Doradus pulsator HD 11231 (Haswell et al., 2019). DMPP-3 is a brown dwarf binary companion in a 507 d orbit, and a hot super-Earth planet in a 6.67 d orbit, around the solar-like primary star HD 42936 (Barnes et al., 2019).

Circumstellar clouds surrounding DMPP planets may, if temperatures are hot enough to sublimate the surface of rocky planets (Rappaport et al., 2012; Perez-Becker & Chiang, 2013), present a tantalising opportunity to perform detailed chemical characterisation on bulk composition material (Fossati et al., 2011). DMPP systems are priority targets for compositional analysis, and will help establish an empirical mass-radius-composition relationship for low-mass planets (Haswell et al., 2019).

1.6 Star-planet interactions

Cuntz et al. (2000) suggested that the activity of the host star may be influenced by interactions with orbiting planets. This phenomenon came to be known as star-planet interactions (SPI), a term which covers tidal effects, stellar radiation, ionising radiation, and magnetic fields.

1.6.1 Mechanisms

In this section, I describe the mechanisms that give effect to tidal and magnetic SPI (SPTI & SPMI respectively).

Tidal SPI

Astronomical objects in close proximity exert mutual gravitational forces. Tidal forces result from the changing gravitational field experienced by the bodies as they rotate around a common centre of mass. This leads to deformation of the bodies, dissipation of energy, and orbital evolution. Energy is deposited in the stellar atmosphere by waves induced by planet-raised tidal bulges (Vidotto, 2017). Modulation of this activity takes place at half the orbital period. For planets within $\sim$ 2 AU of their host star, tidal forces lead to three observable effects: alignment of the spin axes perpendicular to the orbital plane; synchronisation of the rotational and orbital periods of the body; and circularization of the planetary orbit over Gyr time scales (Zahn, 2008; Mathis & Le Poncin-Lafitte, 2009). Synchronisation is the most relevant effect to stellar activity.

Where the planet orbits the star faster than the star rotates (as is generally the case for HJ systems), the planet raises a tidal bulge on the star which lags behind the planetary orbit. The tug of the planet on this bulge raises a tidal torque which pulls the star into a faster rotation—spinning the star up (Figure 1.17). Angular momentum is transferred from the planet to star, which leads to a decrease in a in order to conserve angular momentum in the system. The planet spirals slowly towards the star. Conversely, where the rotation period of the star is shorter than the orbital period, the tidal bulge gets pulled back as angular momentum is transferred to the planet. The star spins down, while the planet migrates to a greater a.

Tidal synchronisation has been observed for stellar binaries which (for spin-up) leads to old, active, fast-rotating stars (Stepien, 1995). SPTI should be efficient in massive late-type stars, where convective layers are thin and vulnerable to the influence of planet-induced tides. This may explain the anomalously low observed activity of WASP-18 ($\log(R'_{\rm HK}) \approx -5.15$: Lanza, 2014), a mid F-type star hosting a $\sim 10\ M_{\rm J}$ planet on a 1-day orbit (Pillitteri et al., 2014; Fossati et al., 2018a). Planets may be engulfed by host stars when they evolve off the main sequence (Privitera et al., 2016b,c,a). With expansion on the red giant branch, SPTI between a Jupiter-like planet and the star may trigger planetary migration that transfers orbital angular momentum to the star, and spins it up. These red giants would present higher rotation rates and magnetic activity than otherwise expected.

Tidal interactions may also effect stellar cycles. The well observed F7V star τ Boo has a 2-year magnetic activity cycle (Catala et al., 2006; Fares et al., 2009; Mengel et al., 2016). Tidal interactions between the thin convective envelope of the star and the planet may have sychronised, while the interior radiative core continues rotating at a different rate, creating a shear that results in short magnetic activity cycles. SPTI indirectly contributes to altering stellar magnetic fields.

Magnetic SPI

The magnetic fields generated by dynamo processes in stellar convective zones shape the orbital environment of close-in planets (Donati & Landstreet, 2009; Brun et al.,

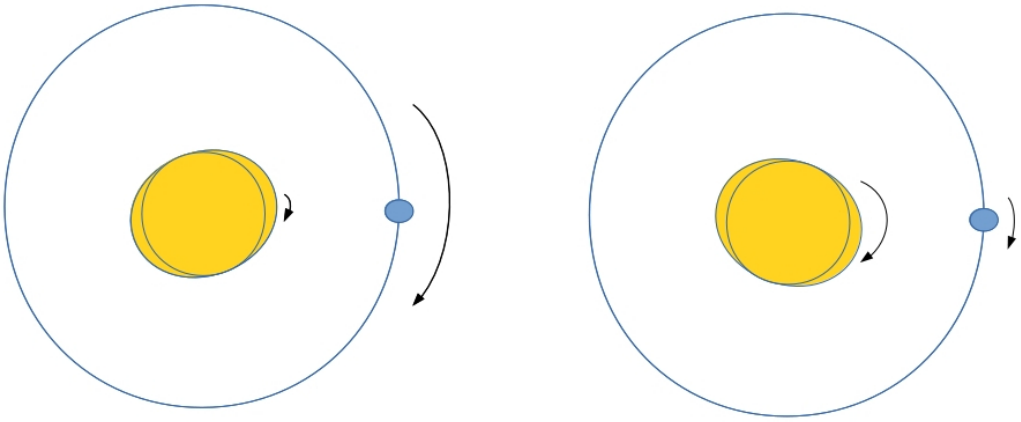

FIGURE 1.17: Tidal SPI interaction. Left $P_{\rm rot} > P_{\rm orb}$. Right $P_{\rm orb} > P_{\rm rot}$. (Poppenhaeger, 2016)

2015a,b). If a planet possesses ionized material, then the bodies will interact magnetically during orbit. Close-in planets that orbit within the stellar Alfvén radius—being the distance within which magnetic energy density is equal to kinetic energy density ($\lesssim R_\star$ or $\lesssim$ 0.1 AU for sun-like stars: Belenkaya et al., 2014)—interact directly with the stellar atmosphere. This interaction may take several forms: magnetic reconnection events between magnetic field lines of a planet and star that accelerate particles along stellar magnetic field lines (Figure 1.18); propagation of Alfvén waves; and through the generation of electron beams that may smash into the stellar corona (Strugarek, 2018). Magnetic interactions can dominate tidal effects where the stellar magnetic field is strong and dynamical tides are not in operation (e.g. in fully convective M dwarfs: Strugarek et al., 2017). These effects can influence stellar rotation and lead to magnetically-driven planet migration (Ahuir et al., 2020). The magnetic torque depends on stellar/planet structure and magnetic field strength, and on orbital separation.

Magnetic interactions 'tickle' the outer layers of the stellar atmosphere, stimulating additional activity (Wright & Miller, 2015). Strugarek et al. (2019) find that, despite being relatively small, Kepler-78b can channel energy flux densities up to a few kW m^{-2} towards Kepler-78 by SPMI, which should be an observable effect. Cohen et al. (2011) model HD 189733b orbiting in and out of a variable Alfvén radius. They find SPI depends on stellar and planet magnetic field orientation and fluctuates during orbit. Theoretical investigations predict magnetic effects, such as hot spots and flares

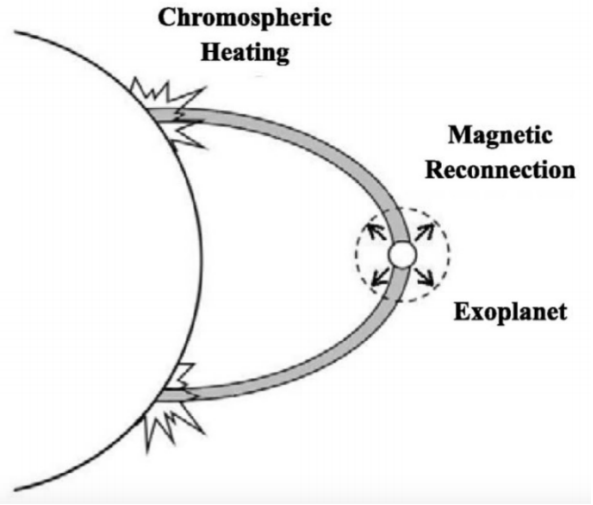

FIGURE 1.18: Illustration of energy channelling and reconnection due to magnetic SPI. (Ip et al., 2004)

(Lanza, 2008; Cohen et al., 2011). Folsom et al. (2020) find that 55 Cancri b orbits inside the Alfvén surface. Planetary magnetic effects can therefore be transmitted to the stellar surface and induce SPI. Cauley et al. (2019a) approximate surface magnetic field values for HJs to be $20 - 120$ G. Large planetary magnetic fields such as these may produce electron-cyclotron maser radio emission, which should be observable by intensive radio monitoring campaigns.

Timing

Tidal and gravitational interactions should display different periodicities. Magnetic SPI will cause one activity enhancement per orbital period, while tidal bulge pairs will induce two activity peaks per planetary orbit period (Cuntz et al., 2000). However, Fischer & Saur (2019) warn that temporal variation of SPI can be more diverse, noting that the synodic and semi-synodic periods may contribute to SPI modulation.

1.6.2 Observational evidence

The search for SPI has been challenging. Efforts have focused on identifying orbitally modulated SPI in individual systems, and systematic SPI in planet-hosting populations.

Individual systems

Evidence of SPI has been observed in a handful of extreme systems, which are characterised by high-mass planets in short period orbits, causing increased interaction energy. Stellar magnetic fields may be large in these extreme systems (Fares et al., 2010). The first reported evidence of SPMI was for HD 179949, a hot ($T_{\rm eff} = 6,200$ K), moderately active ($\log(R'_{\rm HK}) = -4.6$) star displaying elevations in Ca II H & K emission that were synchronised with the orbital period of the planetary companion. (Figure 1.19: Shkolnik et al., 2003, 2005, 2008). Saar et al. (2007) identify potential X-ray synchronization with the planet. However, such sychronisation is not present in all epochs. For example, Scandariato et al. (2013) find no evidence of SPI in X-ray or Ca II H & K lines. Modulation with stellar rotation period is present. This on-off SPI may be attributed to a complex, changing stellar magnetic field configuration (Shkolnik et al., 2008; Fares et al., 2012). Similar emission sychronisation was reported for υ And (Shkolnik et al., 2005, 2008). Subsequent observations found activity to be linked to the stellar rotation period rather than the planetary orbit (Poppenhaeger & Schmitt, 2011).

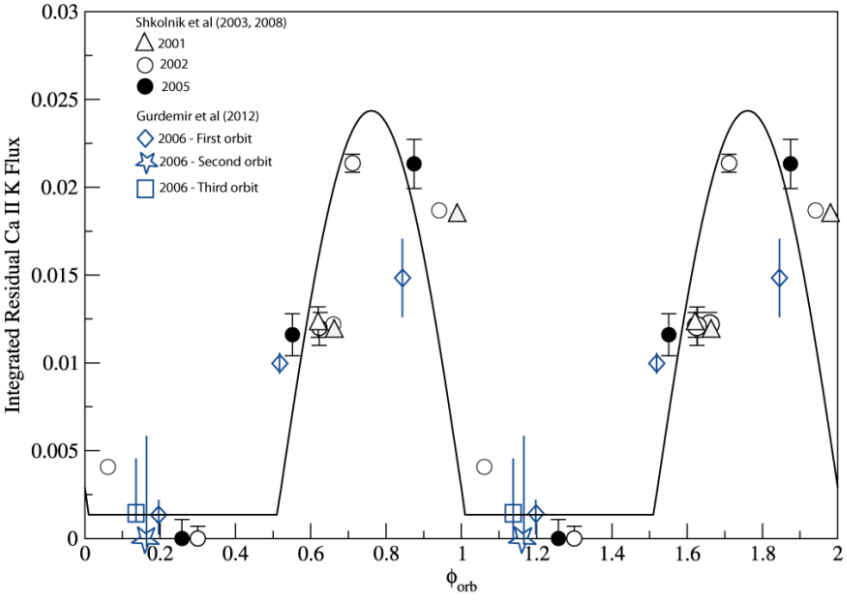

FIGURE 1.19: Phase plot of Ca II K flux for HJ host HD 179949. The sub-planetary point is at $\phi - 0$. Chromospheric activity persistently peaks at orbital phase ~ 0.75. (Shkolnik et al., 2008).

Cool, active star HD 189733 ($T_{\rm eff} = 5000K$ & $\log(R'_{\rm HK}) = -4.5$) is one of the best studied HJs due to its brightness, high activity level, and favourable $(R_{\rm P}/R_\star)^2$ ratio (Bouchy et al., 2005). Flares have been observed in X-ray and UV shortly after a secondary transit at phase $\phi = 0.8$ (Pillitteri et al., 2011). Fares et al. (2010) reported no SPI signal. However, subsequent studies identify an X-ray dark, distant M4V companion, which suggests the system is older than HD 189733's activity indicates (Poppenhaeger et al., 2013). Fast rotation and enhanced X-ray flux may result from stellar spin-up. Pillitteri et al. (2014) suggest that the source of high energy emission may be a trail of in-falling plasma from the planet onto the star, which is most visible just after secondary transit. Maggio et al. (2015) observed elevated X-ray emission during periastron passage and low X-ray emission during apoastron passage, which may be analogous to colliding magnetospheres of young binary stars in eccentric orbits (Getman et al., 2011, 2016). Cauley et al. (2017) demonstrated that Hα line core flux showed abnormal variations around the time of transit. Cauley et al. (2018) use 58 nights of data from multiple epochs containing consecutive night observations to identify modulations in the Ca II K line. A 2.4 d period consistent with the planet's orbital period.

Staab et al. (2017) report enhanced Ca II H & K emission for WASP-43 and WASP-103, which may be attributed to SPI. Fischer & Saur (2019) utilise a semi-analytic parameter study to predict that TRAPPIST 1b and 1c probably generate SPI. They tentatively identify flares occurring in phase with TRAPPIST-1c's synodic period of 9.1 d, and the stellar rotation period of 3.3 d. Kashyap et al. (2008) suggest systems subject to SPI should be X-ray luminous relative to chromospheric activity. However, Miller et al. (2015) find that systems with massive close-in planets are more likely to be active but this may not be caused by direct magnetic interaction. D'Angelo et al. (2019) discuss the possibility of observing SPI in spectral lines during transit with the Colorado Ultraviolet Transit Experiment (CUTE: Fleming et al., 2018).

Population studies

Detecting net activity change within a population of planet hosts has proved challenging. Systematic statistical studies explore whether hot Jupiter hosts have unexpected activity levels. Determining the expected activity level is in itself a challenge that requires a carefully chosen control sample. Poppenhaeger & Schmitt (2011) highlight selection effects in finding hot Jupiters that complicate analysis.

An initial study by Kashyap et al. (2008) analysed X-ray emission from a sample of planet-hosts, finding that stars with small, far-out planets are four times less X-ray active than those with close-in, massive planets. However, several problems with this sample have been identified: it contains brown dwarfs orbiting very active stars (e.g. Cha Hα B orbits a 3 Myr star: Joergens et al., 2010); there is large scatter within the trends (Poppenhaeger et al., 2010); and selection effects and observational biases can explain the trends (Poppenhaeger & Schmitt, 2011). Analysis of X-ray data collected with a more sensitive instrument did not reveal evidence for enhanced activity among hot Jupiter hosts (Poppenhaeger et al., 2010).

Hints of a correlation between stellar X-ray luminosity and planet mass for close-in planets were identified by Scharf (2010). However, Poppenhaeger & Schmitt (2011) found no such relation when using more precise *XMM-Newton* data, identifying published correlations as selection effects. Correlations between X-ray luminosity and SPI proxies (M_P/a^2 and $1/a$) were identified in a sample of 198 FGK MS stars by Miller et al. (2015) but again selection effects were identified. The trend was attributed to a subset of extreme systems with $M_P/a^2 > 450\ M_J\ \mathrm{AU}^{-2}$.

Some studies have had more success in identifying systematic SPI. Poppenhaeger (2015) identify two wide binary systems where the planet host has higher magnetic activity and rotation levels than the companions. The authors suggest tidal interaction between HJs and stars with thick outer convective zones may inhibit stellar spin-down. The Measurements of the Ultraviolet Spectral Characteristics of Low-mass Exoplanetary Systems (MUSCLES: France et al., 2016; Loyd et al., 2016; Youngblood et al., 2016) survey has facilitated the exploration of SPI as a function of emission line formation temperature and led to the tentative detection of SPI as enhanced

transition region activity (France et al., 2016). Expanding on this work, France et al. (2018) used *HST* observations of 71 planet hosts to identify a correlation between FUV activity indices and an SPI proxy (M_P/a). However, once interdependencies between system parameters were accounted for, the correlation was found not to be significant.

Lessons in SPI

Evidence of SPI in individual systems is mixed and seems to be a phenomenon observable at particular epochs. SPI may may be sporadic. Data collected over multiple epochs with consecutive nights covering the orbital period of the planet are likely required to make firm conclusions on orbitally modulated activity (Cauley et al., 2018).

The systematic search for SPI has proved more challenging. Several important lessons have emerged. It is crucial to take into account system age to distinguish between young active stars and older stars that are less active due to effects of magnetic braking (Penev et al., 2012). As it is difficult to estimate the ages of field stars over a Gyr (Soderblom, 2010), approaches for ensuring stars of a similar age are being studied are required. These include studying stars in binary systems where one star hosts a close in exoplanet (e.g. Poppenhaeger, 2015), studying members of the same cluster, or studying stars which are evolutionary similar e.g. solar-like stars (Lorenzo-Oliveira et al., 2018). Systematic observation effects must be mitigated in SPI searches (Poppenhaeger & Schmitt, 2011; Miller et al., 2015; France et al., 2018). Systematic effects also have important consequences for stellar activity. For example, high X-ray and FUV flux radiated by very active stars may lead to higher rates of mass-loss from exoplanets (Sanz-Forcada et al., 2010; Lecavelier des Etangs et al., 2004) and potentially total atmospheric loss (Lopez & Fortney, 2013; Poppenhaeger et al., 2012). Observed activity may not reflect true activity.

1.7 Thesis overview

In this thesis I present the following:

- A homogeneous chromospheric activity survey of 104 close-orbit, transiting planet hosts—the first of its scale and precision in planet activity work;
- Strong evidence for mass loss and enshrouding in close-orbiting systems based on the proportion of the OU-SALT sample that exhibits sub-basal activity compared to field stars;
- A multivariate analysis that provides evidence of star-planet interactions (SPI) with up to $4\,\sigma$ certainty;
- A methodical search for correlations between activity and stellar/planetary variables, with *prima facie* evidence of enshrouding and SPI;
- Analysis that quantifies how absorption in the interstellar medium affects measured activity;
- Constraints on the parameters of stars around which diffuse circumstellar gas discs form;
- Confirmation that WASP-43 has anomalously high activity—supplementing evidence for tidal SPI in this system;
- Evidence of significant, potentially orbitally-modulated activity variation for HD 73256, K2-32, KELT-11, WASP-63, WASP-67, WASP-103 and WASP-104;
- A novel approach to search for magnetic SPI in planet host samples;
- A correction to the largest published activity catalogue that effectively saves the Vaughan-Preston gap;
- A rotation period range for, and constraints on the photometric amplitude of, HD 184960 that may facilitate the detection of a terrestrial companion; and
- A robust, multi-algorithmic approach to identify the rotation periods of low activity stars.

This work provides a blueprint for activity studies of the future.

The thesis is structured as follows: In Chapter 2, photometry and period analysis are employed to identify the rotation period of planet-host candidate HD 184960. Chapters 3 – 6 are dedicated to a novel, high precision activity study we have called 'The OU-SALT Survey' . In Chapter 3, observations made with the Robert Stobie Spectrograph are introduced. The systems observed are explored in Chapter 4. An extensive population study is undertaken in Chapter 5 to search for correlations between stellar activity and stellar/planetary parameters. In Chapter 6, I undertake a statistically rigorous multivariate analysis to search for signals of mass loss and SPI in our unique dataset. Chapter 7 summarises the findings in this thesis and discusses future work.

Chapter 2

The rotation period of HD 184960

This chapter is structured as follows: In Section 2.1, I introduce HD 184960 and explain the importance of identifying its rotation period. I detail our observations of the target with the Las Cumbres Observatory Global Telescope in Section 2.2, as well as the data reduction and photometry I employ on B-band observations in Section 2.3. In Section 2.4, I apply three algorithms to search for period signals in photometry outputs, then fold data onto the most promising candidate periods. This process is repeated for defocused R-band observations in Section 2.5. Results are discussed and used to constrain the parameters of HD 184960 in Section 2.6. A summary and collaboration details are provided in Section 2.7.

2.1 HD 184960

The Dispersed Matter Planet Project (DMPP) has published planet discoveries papers for three systems: DMPP-1, 2 & 3. HD 184960 may become DMPP-4. It is a bright ($V = 5.71$), ~ 300 Myr star of spectral type F7V located 25.6 pc from Earth in the Ursa Major constellation. It has projected rotational velocity $v \sin i \sim 7$ km s^{-1} and a far infra red excess indicative of dust emission (Spangler et al., 2001).

HD 184960 has $\log(R'_{\rm HK}) = -5.24$ (Buccino & Mauas, 2008). This activity level is unexpected given the star's Ursa Major group membership (Montes et al., 2001), which would imply a minimum activity of $\log(R'_{\rm HK}) \sim -4.7$ (King et al., 2003). Observed activity is below the main sequence basal limit. This indicates the absorption of emission flux extrinsic of the star (Haswell et al., 2019; Staab, 2018).

The DMPP team began observing HD 184960 with SOPHIE in 2015. Intensive observations were also made with HARPS-N with a 1-night allocation in July 2016. To date, 154 radial velocities have been collected, revealing an 11-13 $M_\oplus$ planet with orbital period $P_{\rm orb} = 2.673$ d (without HARPS-N data) or 3.266 d (with HARPS-N data). Radial velocities also indicate a possible second low-mas planet with $P_{\rm orb} =$ 0.211 d or 0.266 d.

Isochrone dating suggests that HD 184960 has a rotation period > 10 d. However, given $v \sin i = 7.5^{+0.5}_{-2.0}$ km s^{-1} (L. Fossati, private communication, 2019) and $v \sin i = 2\pi R \sin i / P$, then $P = 2\pi R \sin i / v \sin i = 2\pi \times 1.2 \sin i / 7.5^{+0.5}_{-2.0} = 8.09^{+2.94}_{-0.51} \sin i$ d. So the period could be up to 11.03 d.

A variety of methods may be used to rule out stellar contribution to planetary signals, including searching for correlations between the radial velocity and bisector spans (Queloz et al., 2001). In this chapter however, I use photometry to identify a spot-modulated rotation period (see Section 1.3.4).

2.2 Observations

Photometric time-series observations of HD 184960 were collected using a world-wide network of robotic telescopes known as the Las Cumbres Observatory Global Telescope (LCOGT: Brown et al., 2013). The network includes nine identical 0.4 m diameter telescopes located at several observatories including Teide Observatory on Tenerife (TFN) and Haleakala Observatory in Hawaii (OGG), and a 2 m telescope at Siding Spring Observatory, Australia. The telescopes are centrally controlled to act as a single scientific instrument. The 0.4 m telescopes are supported on a C-ring equatorial mount, with cameras mounted at the Cassegrain focus. Each has a charged couple device (CCD) with 2000 K $\times$ 3000 K pixels (each 0.57 arcseconds square) for a total field of view (FOV) of 19.0 $\times$ 28.5 arcminutes. There is a default 2 $\times$ 2 binning mode. A 14-position filter wheel contains Sloan ugriz primed, Johnson-Cousins V & B, and Pan-STARRS w filters.

The FOV was centred on HD 184960 and frames were obtained through B-filter (761 images) between 17 June and 28 August 2016. The telescopes were defocused by a factor of 0.5 to avoid CCD saturation. A total of 761 exposures of $\sim$ 1 s were collected over 46 observing nights. Between 3 and 63 observations were made each night, with 1-hour cadence between batches of $3 - 9$ observations. Fifteen images were collected by the OGG telescope and the remainder by the TFN telescope.

I planned a new observing strategy in 2017. Observations were made in the R-band, with 6 s exposures and a defocusing factor of 3.0. It was hoped that, despite being less sensitive to high contrast star spots (as contrast within the photosphere decreases with wavelength), images captured through an R-filter would be less sensitive to changes in the sky background at different lunar phases. Four well-spaced visits per night were made with 6 frames taken per visit, over 6 weeks to ensure a long baseline. A greater degree of de-focusing allowed for longer exposures and increased stability. 564 R-filter images were collected.

2.3 Photometry

Figure 2.1 shows a typical image of the field with a scale of 1.16 arcsecs per pixel, and a full-width-half-maximum (FWHM) of the point spread function (PSF) of $\sim 2-4$ pixels. The Las Cumbres Observatory (LCO) pipeline, nicknamed BANZAI, is a general purpose tool designed to reduce data for LCOGT users (McCully et al., 2018). The pipeline includes bad-pixel masking, bias subtraction, dark subtraction, flat field correction, source extraction and an astrometric solution.

I performed aperture photometry using the Image Reduction and Analysis Facility (IRAF) to check and, if possible improve upon, the photometric precision of the LCO pipeline. I used 100 iterations of a centroid algorithm within two 10 000 pixel squares to locate the target and reference stars. To minimise the risk of flux leaking from the defocused stars into the sky annulus, I chose an inner radius of the sky annulus (r) of 20 pixels and an outer radius (R) of 40 pixels, giving an annulus width $R - r = 20$ pixels. Multiple aperture radii of between 4-20 pixels were tested in 2 pixel increments to identify an optimal aperture width.

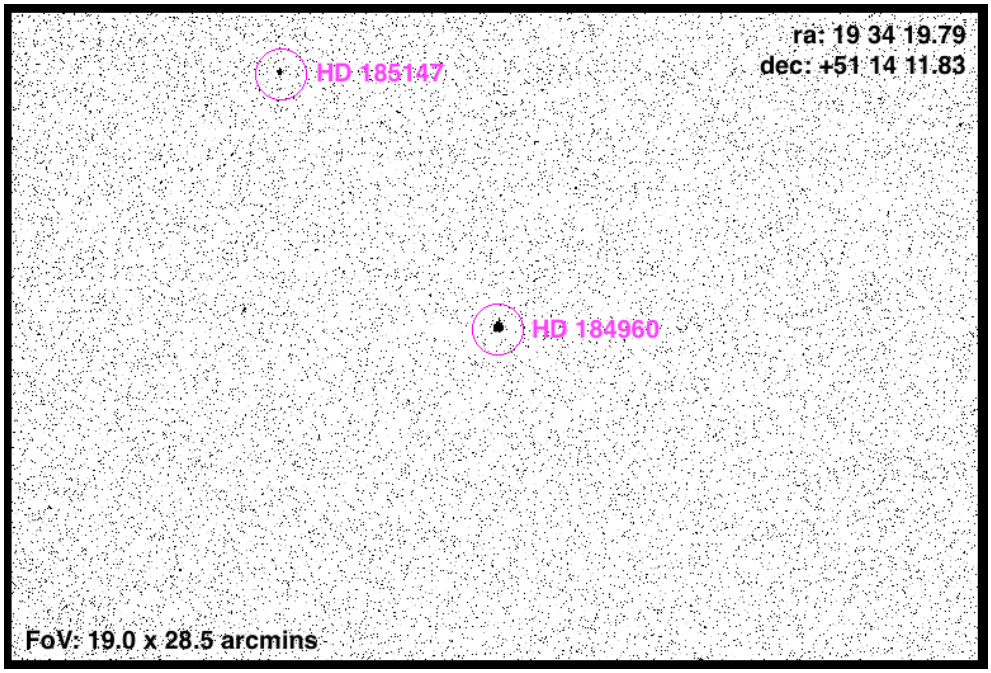

FIGURE 2.1: A typical science frame with inverted colour scale, as collected by LCOGT.

I used IRAF's *apphot* function to extract the following key outputs: the total number of counts in the target aperture less the estimated sky contribution ($c_t(i)$); the total number of counts in the reference aperture less the estimated sky contribution ($c_r(i)$); and the error in each of these values: $\epsilon_t(i)$ and $\epsilon_r(i)$ respectively (which account for photon, sky and digitization noise).

Ideally flux from multiple stars would have been averaged and used as a mean reference for the target. However, HD 185147 (the second brightest star in Figure 2.1) is the only source in the FOV that is bright enough to act as a reference in the B-band. It is of spectral type G5, with visual magnitude V= 7.99. It has very low activity ($\log(R'_{\rm HK}) = -5.21$: Strassmeier et al., 2000) and so is expected have a small photometric amplitude. Any photometric variation present is likely to be on a different timescale to that of HD 184960.

The aim of working with divided counts $d_i = c_t(i)/c_r(i)$ rather than raw data is to cancel out seeing or atmospheric effects that systematically change with time. The propagated error ϵ_i for d_i is calculated:

$$\epsilon_i = d_i.\sqrt{\left(\frac{\epsilon_t(i)}{c_t(i)}\right)^2 + \left(\frac{\epsilon_r(i)}{c_r(i)}\right)^2},$$

where $\epsilon_t(i)$ is the target count error and $\epsilon_r(i)$ is the reference count error. Figure 2.2 shows the Modified Julian Date (MJD: being the Julian Date less 2 400 000.5 days) plotted against divided flux with propagated errors.

The percentage error of d_i ($\epsilon_{\%}$) for each tested aperture was calculated as follows:

$$\epsilon_{\%} = \frac{\sigma_i}{\hat{D}} \times 100,$$

where σ_i is the standard deviation and $\hat{D}$, the inverse variance weighted average of d_i across the time series, was calculated:

$$\hat{D} \equiv \left(\sum_{i=1}^{N} \frac{d_i}{\epsilon_i^2}\right) \Bigg/ \left(\sum_{i=1}^{N} \frac{1}{\epsilon_i^2}\right).$$

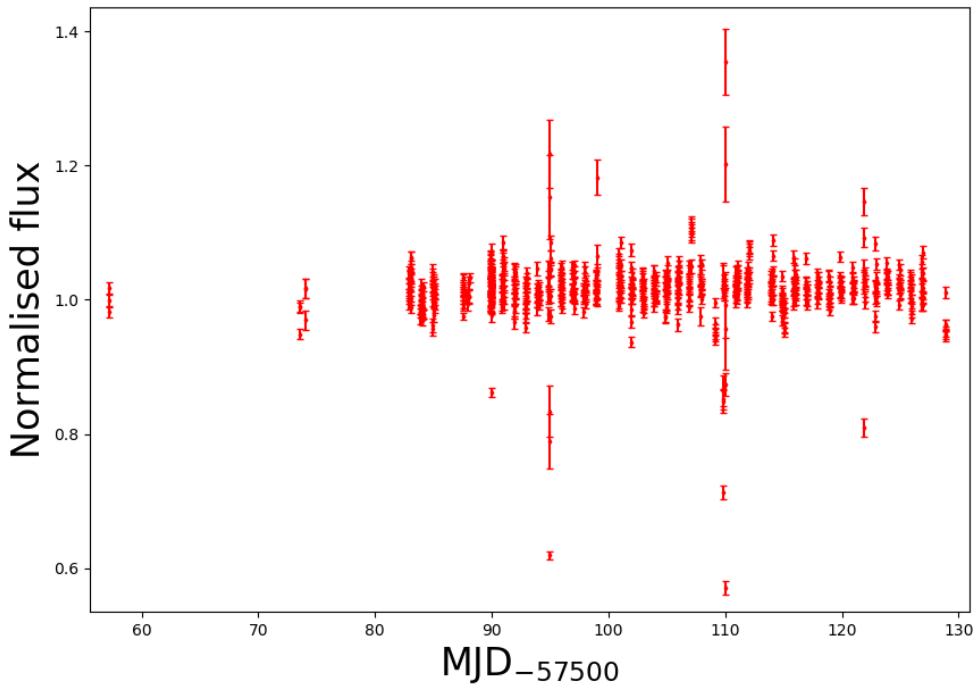

FIGURE 2.2: MJD plotted against divided counts normalised by the mean for B-band observations. Propagated errors are shown.

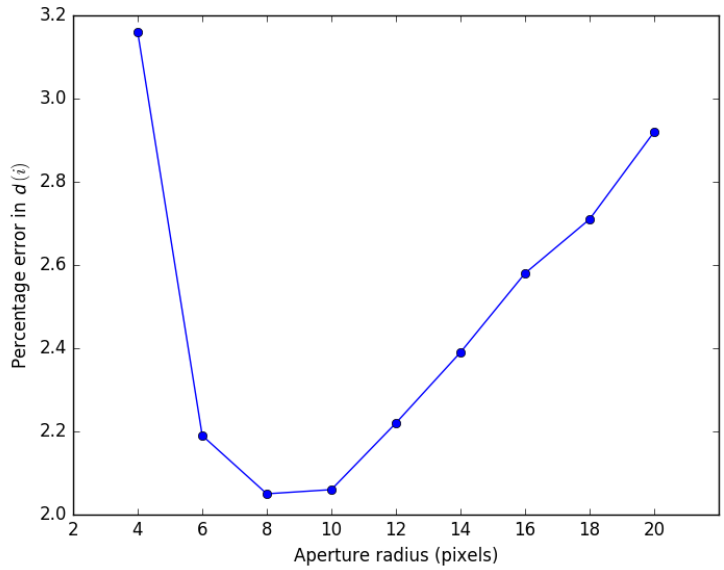

FIGURE 2.3: Aperture radii used for IRAF photometry plotted against percentage error in d_i for the B-filter time series. Note the characteristic shape of plot. Ap8 and Ap10 provided the lowest percentage error in d_i, and highest photometric precision.

Figure 2.3 shows the percentage errors for each aperture after an 11 σ clip was applied to exclude unreliable observations. Aperture radii of 8 and 10 pixels (Ap8 & Ap10) provided the highest precision photometry.

2.3.1 Quality checks

Centroid tracking

I used a centroid algorithm to identify stellar centres. The coordinates of the target and reference star centres were identified in a typical frame: x_{ti} and y_{ti} for the target and x_{ri} and y_{ri} for the reference star (the 'Initial Coordinates'). The algorithm then searched within a 1000×1000 pixel box to find the peak intensity weighted means of the profiles in x and y of each star. It identified x_{tf} and y_{tf} for the target star and x_{rf} and y_{rf} for the reference star (the 'Final Coordinates'). Assuming negligible rotation of the FOV between frames, it was expected that the displacement between the Starting Coordinates and Final Coordinates should be the same for both the target and reference stars: $(x_{tf} - x_{ti}) = (x_{rf} - x_{ri})$ and $(y_{tf} - y_{ti}) = (y_{rf} - y_{ri})$. For 96% of images $-0.6 < [(x_{tf} - x_{ti}) - (x_{rf} - x_{ri})] < 1.4$ pix and $-1 < [(y_{tf} - y_{ti}) - (y_{rf} - y_{ri})] < 1$ pix. Centroid tracking therefore matches for the target and reference in the majority of frames.

Systematic offsets

Figure 2.4 shows the residuals between each x and y coordinate and coordinate mean ($x - \bar{x}$ and $y - \bar{y}$ respectively) with a fitted Gaussian curve. For the target, x and y residuals are well centred with means of 4.91×10^{-14} and -9.412×10^{-14} pix. There is a larger spread of y residuals with standard deviation $\sigma_y = 12.57$ pix compared to $\sigma_x = 5.44$ pix. Outliers are evident for the reference star, with a group of y residuals exceeding 150 pix. These outlying points are caused by poor sky transparency and seeing, and are removed from subsequent analysis. The x residuals are well centred, with a mean of -2.3×10^{-14} pix. The y distribution appears slightly offset but the outliers shift the distribution mean back close to zero (1.1×10^{-13} pix). The standard deviations are $\sigma_x = 18.97$ pix and $\sigma_y = 48.79$ pix, where the latter is affected by the

outlying group. When the outlying points are removed, the values become $\sigma_x = 5.59$ pix and $\sigma_y = 12.6$ pix, which are similar to the target. The centroid algorithm performs less consistently on the fainter reference star but no strong systematic offset is evident.

Airmass effects

As HD 184960 and HD 185147 are different spectral types, airmass effects could introduce spurious variability. Pearson tests (Pearson & Karl, 1895) were performed to check for linear correlations between airmass and divided flux. Correlation coefficient (r) values of -1 and 1 indicate perfect anti- and positive correlation, while zero equates to no correlation. For Ap10, $r = 0.058$ with a p-value of 0.12. Similar values are found for Ap8. For the LCOGT pipeline, $r = -0.063$ with $p = 0.088$. The r-values are close to zero, indicating weak correlations. As $p > 0.05$, the null hypothesis (that airmass and divided counts are not significantly correlated) is accepted.

2.3.2 LCO pipeline

For LCO pipeline-extracted outputs, the weighted mean of the d_i array is 11.62, which has a considerably larger percentage error (12.1 %) than Ap8 and Ap10. This is partly caused by 16 points with $d_i < 10$ that have propagated error $< 11\,\sigma$, so are not excluded by the error clip. On clipping these points (the presence of which may suggest some irregularities in the LCOGT reduction), the percentage error of the remaining 727 points drops to 2.6 %.

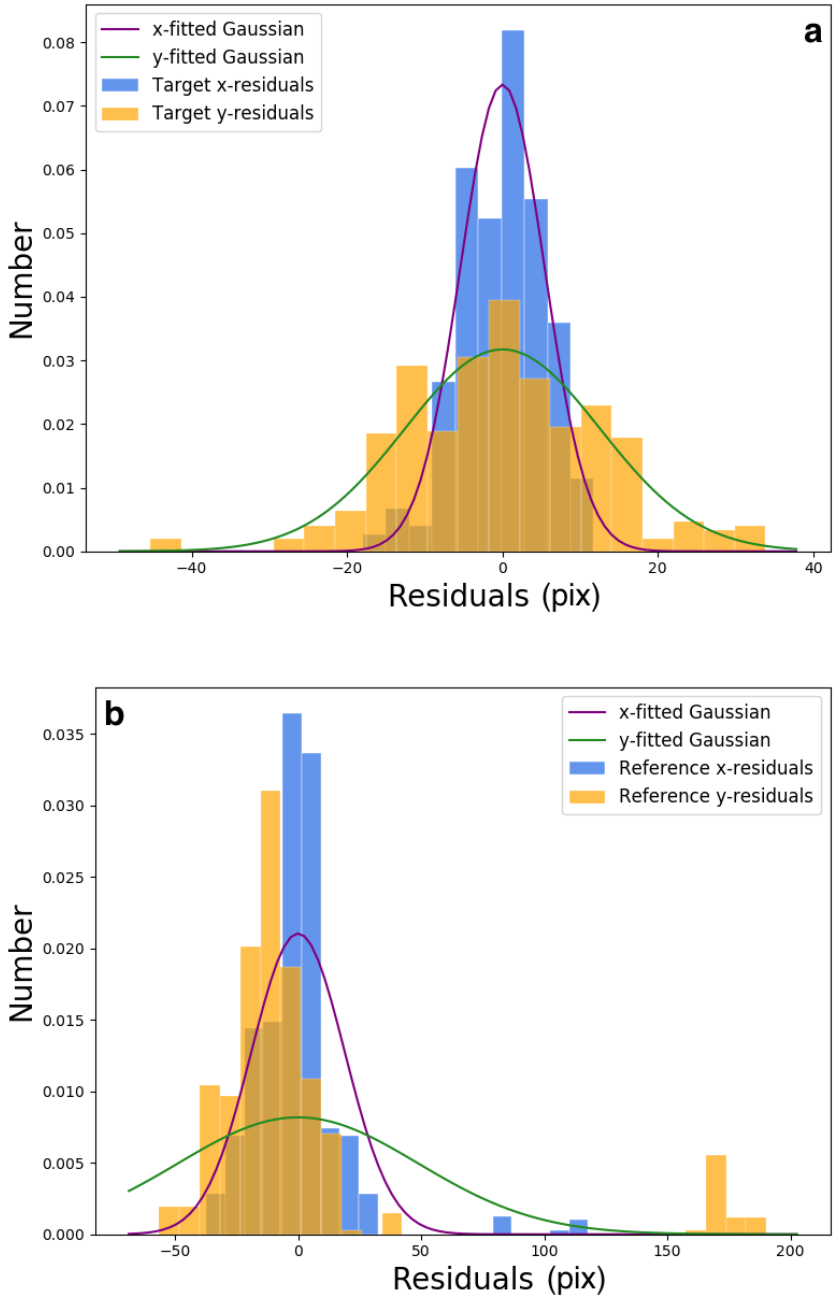

FIGURE 2.4: (a) Residual distributions for target x and y coordinates, with best-fit Gaussian curves plotted. (b) As per top panel but for the reference star.

2.4 Period search

I used three algorithms to search for HD 184960's rotation period: the Lomb-Scargle (LS) periodogram, the string-length (SL) method; and phase dispersion minimisation (PDM).

2.4.1 Lomb-Scargle periodogram

The LS periodogram (Lomb, 1976; Scargle, 1982) is a tool for detecting periodic signals in unevenly-sampled data. It involves fitting a sinusoidal model at a range of frequencies in order to compute a power. Larger power reflects a better fit. The algorithm, which is motivated by Fourier transforms, is commonly used in astronomy (VanderPlas, 2018). I used the *Time series* module in *Astropy* (Robitaille et al., 2013; Price-Whelan et al., 2018) to implement LS period searches. I trialled both 'slow' (Lomb, 1976; Scargle, 1982) and 'fast' (Press & Rybicki, 1989) implementations (where the latter uses a fast Fourier transform to reduce the number of required operations), finding that they identify equivalent periods.

A key decision when implementing the LS algorithm is how many frequencies to compute. Choosing a coarse frequency grid may mean periodogram peaks are missed, while a fine grid is computationally expensive. I specified a period range of $0.1 - 10$ days in 10000 steps to avoid excessive computation of extremely small periods, and achieve an even sampling of the period range.

To consider the significance of identified peaks, I calculated the false alarm probability (FAP)—being the probability of measuring a peak in data consisting only of non-periodic, Gaussian noise. A number of different calculation methods are available: Baluev (2008) calculates an upper-limit to the alias-free probability; Davies (1987) also provides an upper limit; a 'naive' approach that approximates probability based on an estimated effective number of independent frequencies; and boostrap that resamples the input data by replacement to approximate probability (Suveges, 2012, see Chapter 6 for a description of the boostrap method). Bootstrap is the most powerful but computationally expensive approach.

The periodograms for Ap8, Ap10 and the LCOGT pipeline are presented in Figure 2.5, with the strongest periods and corresponding powers noted in Table 2.1. For Ap10, I show the FAP = 0.1 % level of all four approaches, and FAP = 1 % and 10 % for the boostrap method in the lower panel of Figure 2.5. FAPs are placed in the context of LS powers and bootstrapped peak powers in Figure 2.6. The LCO pipeline produces the most powerful signals, with LS power > 0.06. The strongest signal is at 1.24 d (FAP = 1.27×10^{-8}). There are several other strong sub two-day peaks and a notable signal at 5.11 d (FAP = 1.28×10^{-6}). The strongest signal for Ap8 is at 6.07 d (FAP = 5.36×10^{-5}), with several sub two-day peaks also present. The most powerful Ap10 signals are 5.00 d (FAP = 2.18×10^{-3}) and 6.07 d (FAP = 2.82×10^{-2}). The Ap8 and Ap10 sub two-day peaks are less significant relative to the $\sim$ 5 d signal than in the LCOGT periodogram. Photometry performed with IRAF appears to produce fewer spurious peaks than the LCO pipeline.

The window function shown in Figure 2.5 indicates periodicities that are due to data sampling alone. Day and sub-day alias signals are present, with the function also strengthening at long periods. The window function power is low at intermediate periods, suggesting LS signals can be considered genuine in this range.

Taking the most powerful $>$ 2 d peak for each reduction method (5.11 d, 6.07 d and 5.00 d respectively), the FAPs indicate that—under the assumption that there is no periodic signal in the data—we will observe a peak this high or higher approximately 1.28×10^{-4} %, 5.36×10^{-3} % and 2.18 % of the time. This provides a strong indication that the periods identified are significant. The LS power of these signals (which may be in the range $0 - 1$) are however low. The estimated FAPs are based solely on calculated power values. This means weak spurious signals could appear significant in these low-power distributions. Moreover, the $\sim$ 5 and 6 d signals could be artefacts related the Earth's rotation period. Further proof is required that the signals are significant.

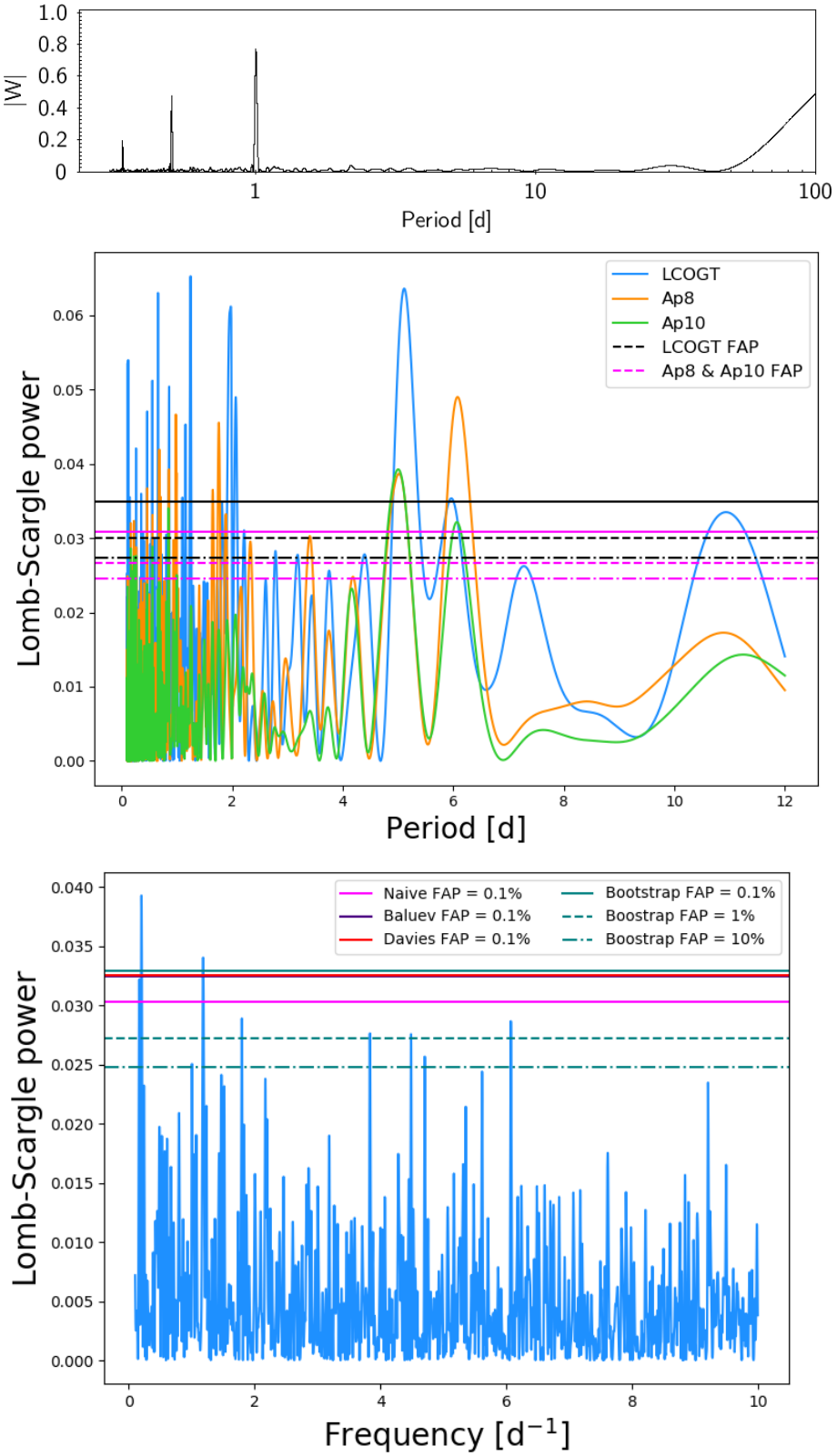

FIGURE 2.5: Top panel: The window function for B-filter observations showing periodicities that are due to data sampling alone. Note the logarithmic x-axis scale. Middle panel: LCOGT, Ap8 and Ap10 periodograms. Bootstrapped 0.1% (solid line), 1% (dashed line) and 10% (dashdot line) false alarm probability (FAP) levels are shown for LCOGT (black), and Ap8 and Ap10 (pink) periodograms. Bottom panel: Frequency plot for Ap10. FAP levels shown are calculated with different algorithms as detailed in the legend.

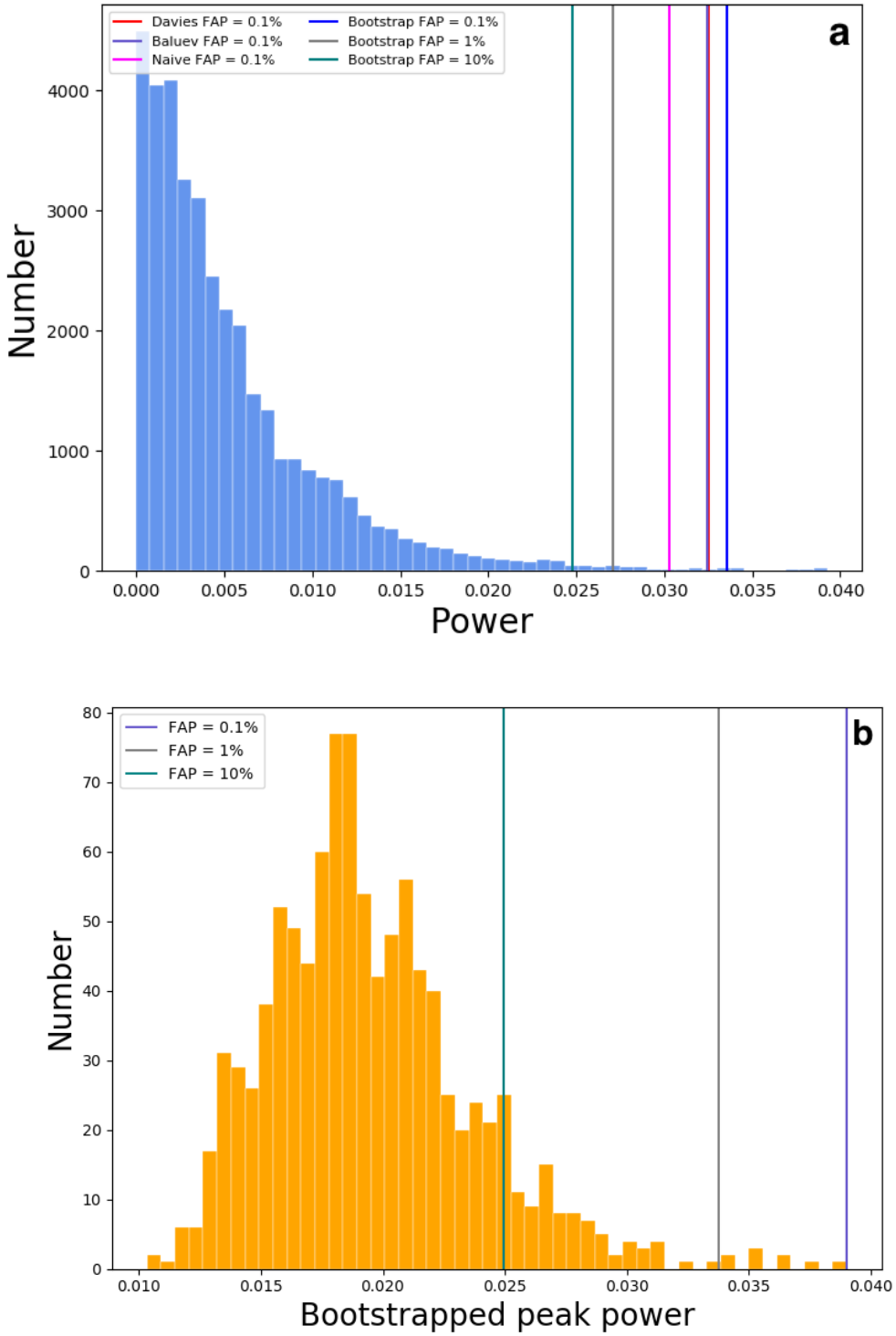

FIGURE 2.6: (a) The power distribution for B-filter Ap10 IRAF-reduced data with FAP levels indicated as described by the legend. (b) The peak LS power for each Ap10 bootstrapped sample, with corresponding FAP levels indicated.

Dataset	Period [d]	LS power	FAP
LCOGT	1.24	0.073	1.27e − 8
	1.94	0.062	7.76e − 7
	5.11	0.061	1.28e − 6
	0.66	0.058	3.67e − 6
	6.01	0.036	9.69e − 3
	11.11	0.034	1.73e − 2
Ap8	6.07	0.049	5.36e − 5
	0.98	0.047	1.30e − 4
	1.76	0.046	1.95e − 4
	0.68	0.042	7.48e − 4
	0.85	0.039	1.99e − 3
	5.01	0.039	2.47e − 3
Ap10	5.00	0.039	2.18e − 3
	0.85	0.034	1.46e − 2
	6.07	0.032	2.82e − 2
	0.56	0.029	8.87e − 2
	0.16	0.029	9.59e − 2

TABLE 2.1: LS periods, powers and FAPs for Ap8, Ap10 and LCOGT B-band reductions.

2.4.2 String-length method

The string-length (SL) method (Lafler & Kinman, 1965; Dworetsky, 1983; Clarke, 2002) searches for periodic signals in time-series data. It is non-parametric and, unlike the LS algorithm, does not assume that a particular (sinusoidal) function describes the data. The method involves folding data onto a number of trial periods, then calculating distances between consecutive data points. The total length of the distances between all points—the string length—is at its minimum when folding produces maximum order.

To find the string length, I first scaled the divided counts:

$$\tilde{d}_i = \frac{(d_i - d_{\min})}{2\,(d_{\max} - d_{\min}) - 0.25'} \tag{2.1}$$

where $d_{\min}$ and $d_{\max}$ are the minimum and maximum divided counts (Dworetsky, 1983).

The sum of the lengths of line segments joining consecutive points in a phase plot (d_i, ϕ_i) were calculated:

$$\mathrm{SL} = \sum_{i=1}^{n-1} \left[\left(\tilde{d}_i - \tilde{d}_{i-1}\right)^2 + \left(\phi_i - \phi_{i-1}\right)^2 \right]^{1/2} + \left[\left(\tilde{d}_1 - \tilde{d}_n\right)^2 + \left(\phi_1 - \phi_n + 1\right)^2 \right]^{1/2}, \quad (2.2)$$

where n is the number of measurements (Burke et al., 1970).

I used the *PyAstronomy* implementation of the Dworetsky (1983) SL method with 1000 trial periods evenly distributed between $0.1 - 10$ d. Calculated SLs are presented in Figure 2.7. The five shortest SLs and corresponding periods are detailed in Table 2.2, along with the residual difference between the mean and calculated SL values (μ_{SL} - SL) expressed in units of standard deviation (σ). These so-called 'Z-scores' were converted to p-values by treating them as a one-tailed hypothesis (i.e. the signals were assumed to be lower than the mean).

The strongest Ap10 signal is 4.18 d ($p = 0.003$). The same period is identified in Ap8 data with $p = 0.005$. A 6.96 d ($p = 0.009$ and 0.007) signal is identified by both Ap8 and Ap10, with $p = 0.009$ and 0.007 respectively. LCOGT periods have lower significance, with the strongest period 7.85 d ($p = 0.007$). Periods ~ 5 d ($p = 0.015$ and 0.007) are identified in both LCOGT and Ap10 datasets.

The SL method produces multiple minima, the significance of which are difficult to assess. It was desirable to employ a second non-parametric method to distinguish between genuine and false peaks.

2.4.3 Phase dispersion minimization

Phase dispersion minimization (PDM: Stellingwerf, 1978) identifies periods that minimize the variance with respect to a mean curve that is defined by binned sub-sets of the measured data. The technique involves folding data onto multiple trial periods, binning it, and computing the variance of the amplitude within each bin. Least squares fitting is used to identify the best period. Unlike Fourier-based analysis, the mean curve is calculated directly from the data rather than by fitting a sinusoid.

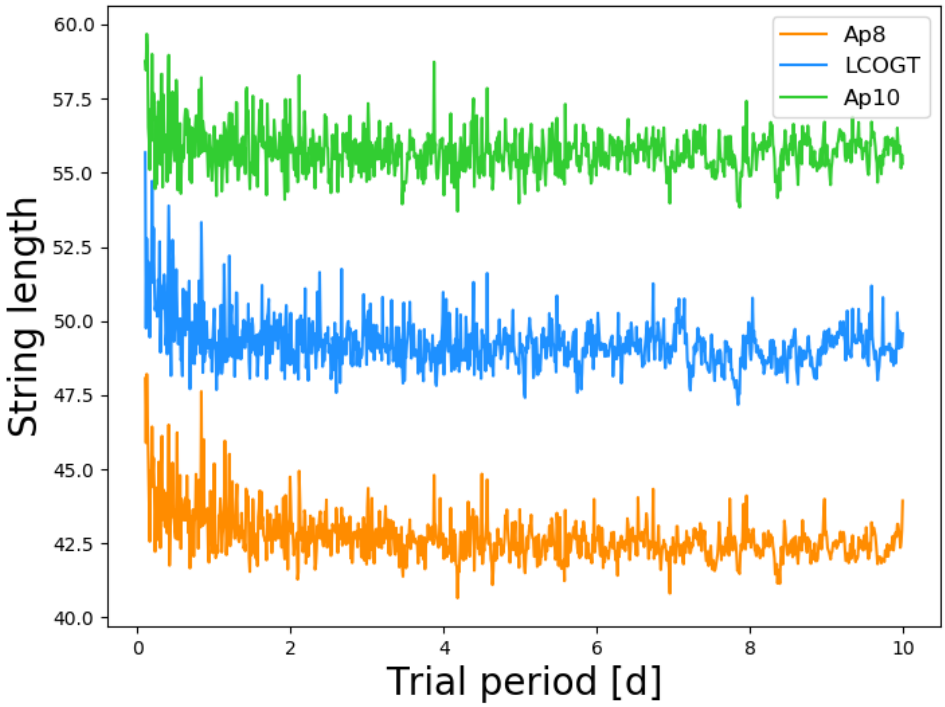

FIGURE 2.7: Trial periods plotted against string length for Ap8, Ap10 and LCOGT data. The LCOGT curve has been shifted down the y axis by 5 for clarity.

Dataset	Period [d]	SL	(μ_{SL} - SL) [σ]	p-value
LCOGT	7.85	51.18	2.45	0.007
	5.07	51.41	2.17	0.015
	2.60	51.58	1.97	0.024
	5.75	51.58	1.97	0.024
	7.22	51.66	1.88	0.030
Ap8	4.18	40.66	2.56	0.005
	6.96	40.81	2.37	0.009
	4.64	41.10	2.01	0.022
	8.40	41.14	1.95	0.026
	5.58	41.23	1.84	0.033
Ap10	4.18	53.70	2.84	0.003
	7.87	53.83	2.65	0.004
	3.50	53.94	2.50	0.006
	4.99	53.97	2.46	0.007
	6.96	53.97	2.46	0.007

TABLE 2.2: String-length results: trial periods with lowest SL values; residuals between the mean and calculated SL in σ; and corresponding p-values.

The sum of binned variances was divided by the sum of total variances of the data to define the following test statistic:

$$\Theta_{\mathrm{PDM}}(\nu) = \frac{s^2}{\sigma^2}. \tag{2.3}$$

The variance of d_i is:

$$\sigma^2 = \frac{\sum \left(d_i - \bar{d}\right)^2}{N-1}, \tag{2.4}$$

where $\bar{d}$ is the mean of d_i, and N is the number of observations. The variance of M binned subsets is:

$$s^2 = \frac{\sum \left(n_j - 1\right) s_j^2}{\sum n_j - M}, \tag{2.5}$$

where s_j^2 is the variance of each sub-set, and n_j is the number of measurements per bin. Test frequencies not present in the data will return $\Theta_{\mathrm{PDM}} \approx 1.0$. The binned variance drops below the overall variance for possible periods and $\Theta < 1.0$. The trial period that produces the lowest Θ_{PDM} value minimizes variance with respect to the mean light curve.

I used the *PyAstronomy* implementation *py(Timing)*, with 10 equidistant phase bins, to carry out PDM. Partitioning data into equal bins can be disadvantageous. For example, if the number of bins is large or data points sparse, bins may contain few data points. I therefore also employed a 'bin-cover' structure—where each bin was phase-shifted by a factor of $1/M \times c$ (where c is a constant) with respect to the previous bin—to ensure data points were covered by multiple bins. I used three cover bins.

The results are plotted in Figure 2.8 for Ap8, Ap10 and LCOGT reductions. In general, searches with no cover bins find lower Θ_{PDM} values than searches with cover bins. This suggests that the fixed bins are well-filled and that using cover bins acts to dilute periodic signals. Only fixed bins were used for further analysis.

Table 2.4 details the trial periods for the five lowest $\Theta_{\rm PDM}$ values in each dataset. I express the significance of these periods as chi-squared (χ^2) p-values. The most statistically significant signal is at 6.25 d in the LCOGT pipeline, which has $p = 0.014$: there is a 1.4% probability that the signal is attributable to random Gaussian noise. Ap8 has 5.00 d ($p = 0.055$) and 6.25 ($p = 0.056$) peaks. For Ap10, periods of 1.67 d ($p = 0.046$ and 5.00 d ($p = 0.062$) are apparent. Again, no outstanding candidate signal is identified.

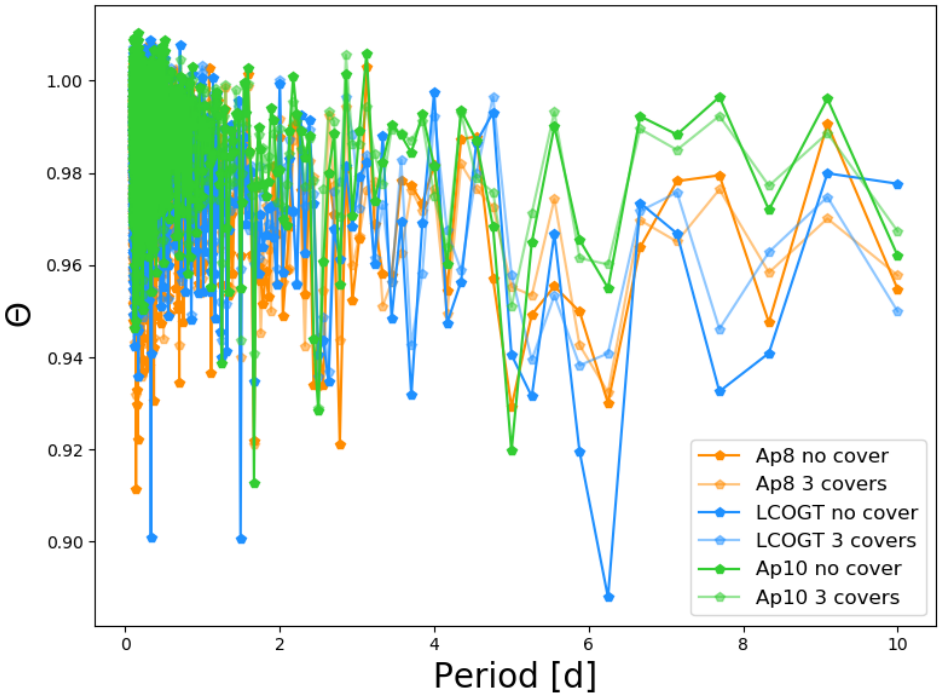

FIGURE 2.8: A plot of trial period and Θ, being the statistic generated by PDM for Ap8, Ap10 and IRAF-reduced, B-filter data. Period searches with and without cover bins are shown.

2.4.4 Phase-folded light curves

I folded B-filter data on to the most promising periods identified above, binned them into 20 bins, and plotted the phase-folded light curves in Figure 2.9. The scatter in each bin is used to estimate uncertainties, which are generally smaller than the marker size. As the 5.01 d period is almost an integer, divided count values are clumped into distinct groups. The data appears to split above and below the best-fit

Dataset	Period [d]	Θ_{PDM}	p-value
LCOGT	6.25	0.888	0.014
	1.49	0.901	0.027
	0.33	0.901	0.027
	5.88	0.920	0.062
	5.26	0.932	0.096
Ap8	0.14	0.911	0.024
	2.78	0.921	0.039
	1.67	0.922	0.040
	5.00	0.929	0.055
	6.25	0.930	0.056
Ap10	1.67	0.913	0.046
	5.00	0.920	0.062
	2.50	0.929	0.087
	1.30	0.939	0.122
	2.44	0.944	0.143

TABLE 2.3: PDM results: trial periods with lowest Θ_{PDM} values, and corresponding p-values.

sinusoid, with both subsets of binned data displaying a degree of sinusoidal variation. Folding the data onto a 4.18 d period produces binned points that are close to the line of best fit but do not appear sinusoidal. The 5.13 d binned points produce the most convincingly sinusoidal signal but the fitted sinusoid is out of phase. The 6.25 d period identified by the PDM method on the LCOGT pipeline data, which is one of the strongest signals, does not produce an obviously sinusoidal curve.

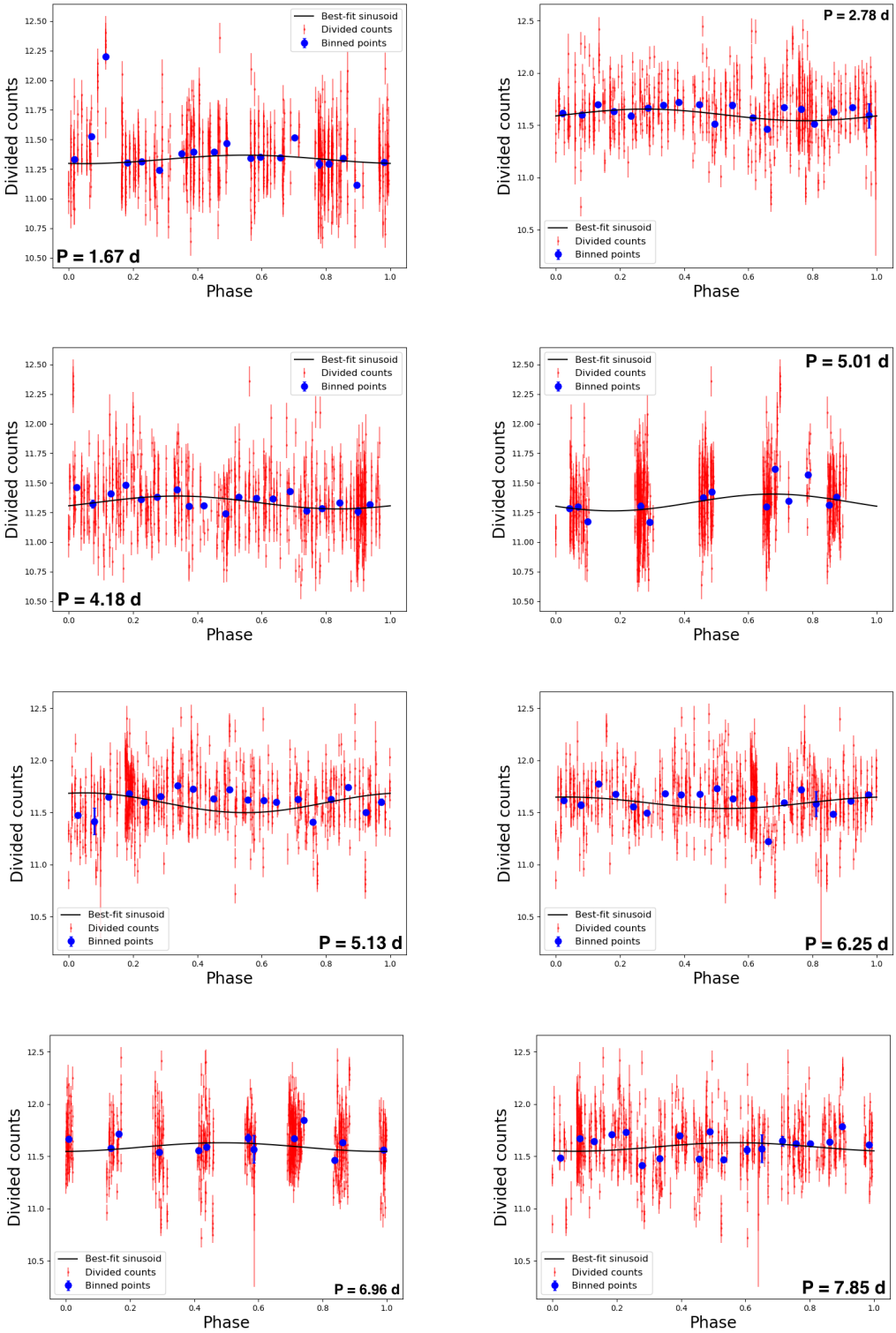

FIGURE 2.9: Phase-folded light curves for candidate rotation periods. Divided counts are shown in red, and are divided into 20 bins (shown in blue). Scatter in each bin is used to estimate uncertainties, which are generally smaller than the marker size. Best fit LS sinusoids are plotted.

2.5 R-band observations

As the B-band period search was inconclusive, the target was re-visited in the R-band using a higher defocusing factor and longer exposures. As sources appeared donut-like in the defocused images, it was not appropriate to use a centroid algorithm to identify stellar centres (Southworth et al., 2009; Basturk et al., 2019). Instead, the position of each star was identified in a typical image, then images were vertically stacked to ensure the stellar centres aligned. This process, and subsequent photometry, was carried out in AstroImageJ (Collins et al., 2017).

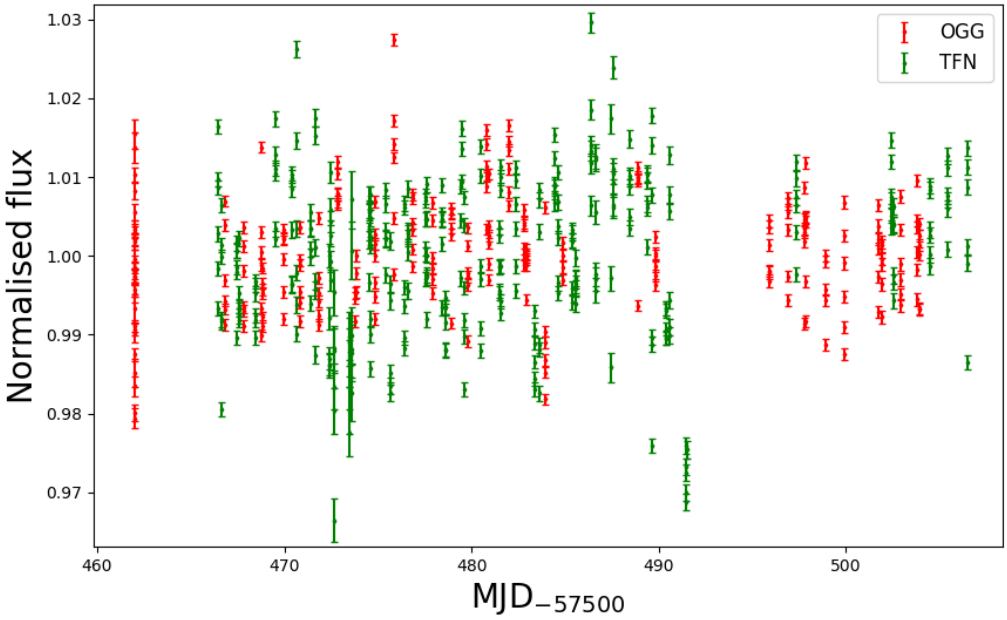

FIGURE 2.10: MJD plotted against normalised divided counts for the majority of LCOGT R-filter observations, with propagated errors. OGG observations are shown in red and TFN in green, with each set normalised by its mean.

2.5.1 Correcting for offsets

LCOGT observations are made using several telescopes at multiple sites. While 98% of B-filter observations were made by a single telescope at TFN, R-filter observations were split between pairs of telescope at the TFN (317 frames) and OGG (247 frames) sites (see Figure 2.10). As image stacking was employed to locate stars, correcting for offsets between telescopes was crucial. No information was available from LCOGT

on systematic effects between the telescopes. An offset between telescopes at OGG was manually corrected by aligning mean values for each. No offset was evident at TFN. Photometry was performed for each site separately, with separate and combined datasets analysed.

2.5.2 Period searches

LS periodgrams

Figure 2.11 shows the LS periodgram for TFN, OGG, combined (TFN & OGG), and the LCOGT pipeline. The 0.1% FAPs are shown for each dataset, and 1% and 10% FAPs also shown for TFN and OGG. A strong 1 d signal, likely caused by Earth's rotation, is seen in all data sub-sets. Several stronger peaks are evident, particularly in the OGG sub-set at 3.12 d (FAP $= 6.33 \times 10^{-3}$) and 5.29 d (FAP $= 2.31 \times 10^{-2}$). In the TFN subset, peaks at 1.64 d (FAP $= 6.69 \times 10^{-3}$) and 8.47 d (FAP $= 6.68 \times 10^{-3}$) have lower power than OGG peaks. LS power is lower again in the combined data set, with peaks at 3.96 d (FAP $= 2.08 \times 10^{-3}$) and a 7.98 d (FAP $= 1.66 \times 10^{-2}$). The LCOGT pipeline-reduced data has weak LS power, which is expected given the LCO pipeline is not calibrated for defocused images. Like B-filter data, the window function indicates alias signals occur at 1-day (though with lower power) and longer periods but are not present at intermediate periods.

SL method

Figure 2.12 shows a plot of trial periods against SL for R-filter data. It is apparent that smaller datasets have shorter SLs. The trial periods with the shortest SLs are presented in Table 2.5. A 4.98 d ($p = 0.058$) period is identified in the combined subset. The stand-out OGG period is 5.55 d ($p = 0.012$), while a 5.6 d ($p = 0.041$) period is present in the TFN data. Periods of 2.67 d ($p = 0.017$) and 6.67 d ($p = 0.035$) are also identified in the TFN sub-set.

Dataset	Period [d]	LS power	FAP
LCOGT-R	0.99	0.210	$1.19e-22$
	0.05	0.099	$1.39e-8$
	0.19	0.070	$3.63e-5$
	0.12	0.069	$3.94e-5$
	4.00	0.033	$3.30e-1$
OGG	0.11	0.176	$8.23e-8$
	3.12	0.094	$6.33e-3$
	5.29	0.084	$2.31e-2$
	1.49	0.075	$6.85e-2$
	7.86	0.065	$2.12e-1$
TFN	0.12	0.133	$6.58e-7$
	1.64	0.081	$6.69e-3$
	8.47	0.079	$6.68e-3$
	2.54	0.066	$5.63e-2$
	3.96	0.062	$9.62e-2$
Combined	0.23	0.039	$2.89e-29$
	0.50	0.121	$8.13e-13$
	3.96	0.048	$2.08e-3$
	7.98	0.041	$1.66e-2$
	1.91	0.039	$2.75e-2$

TABLE 2.4: LS periods, powers and FAPs for Ap8, Ap10 and LCOGT R-band reductions.

Dataset	Period [d]	SL	(μ_{SL} - SL) [σ]	p-value
OGG	9.29	13.15	2.45	0.007
	5.55	13.21	2.25	0.012
	9.29	13.23	2.19	0.014
	0.99	13.25	2.12	0.017
	0.30	13.25	2.12	0.017
TFN	2.67	20.35	2.12	0.017
	9.19	20.41	2.02	0.022
	6.67	20.53	1.81	0.035
	5.60	20.58	1.74	0.041
	4.24	20.59	1.71	0.044
Combined	9.97	30.18	1.58	0.057
	4.98	30.19	1.57	0.058
	6.67	30.35	1.43	0.076
	4.02	30.45	1.34	0.090
	5.48.	30.56	1.24	0.108

TABLE 2.5: R-band SL results: trial periods with lowest SL values, and corresponding residuals between the mean and calculated SL in σ.

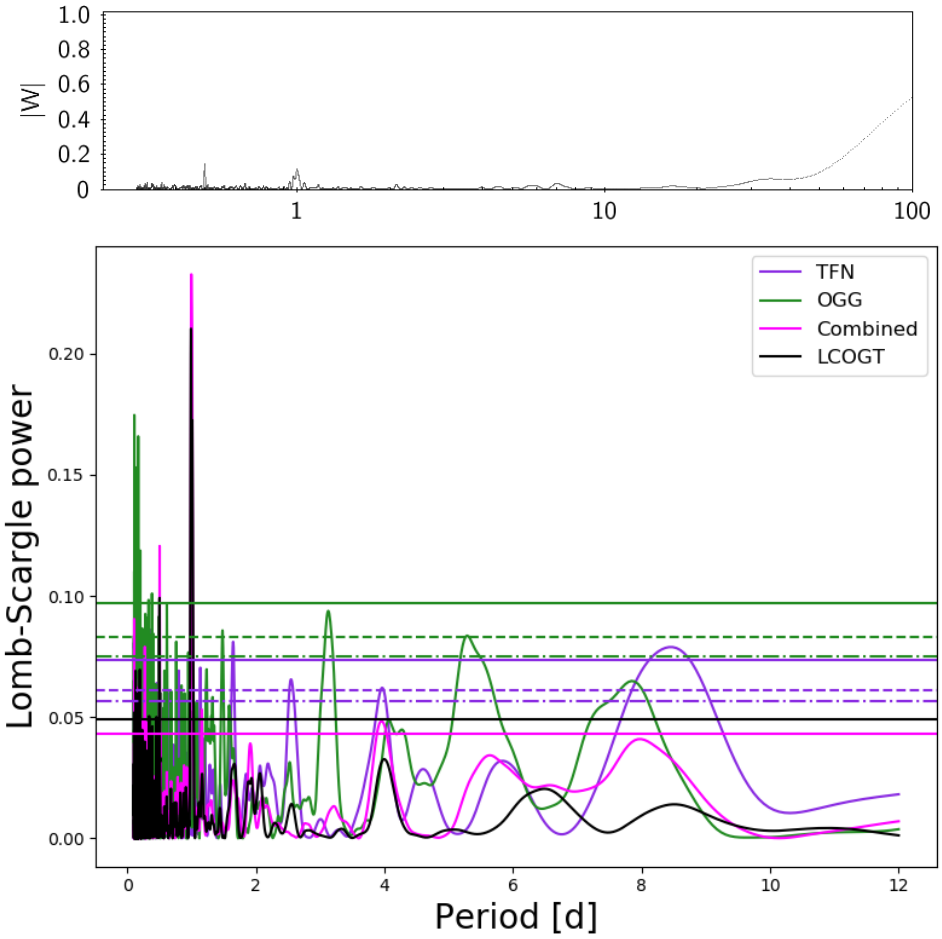

FIGURE 2.11: Top panel: The window function for B-filter observations showing periodicities that are due to data sampling alone. Note the logarithmic x-axis scale. Bottom panel: Periodograms for R-filter sub-sets as indicated in the legend. The 0.1% FAPs are shown for each periodogram, and 1% and 10% FAPs are also shown for TFN and OGG sub-sets.

PDM

Figure 2.13 shows the Θ_{PDM} values for the TFN, OGG and combined datasets, while Table 2.6 details the trial periods with the lowest Θ_{PDM} values. Where multiple signals are found in close temporal proximity, only the strongest signal is described. The OGG shows minimum Θ_{PDM} values at 1.61 d ($p = 0.067$) and 6.25 d ($p = 0.091$), while there are minimum values at 2.56 d ($p = 0.023$) and 5.56 d ($p = 0.042$) in the TFN data. The most notable period in the combined data is at 4.00 d ($p = 1.7 \times 10^{-3}$), though with a strong signal at 1.0 d, this could be caused by an artefact related to Earth's rotation period. The p-values calculated are low, especially for the combined sub-set, suggesting the periods identified are significant. There remains, however, no stand-out candidate signal identified by all three algorithms.

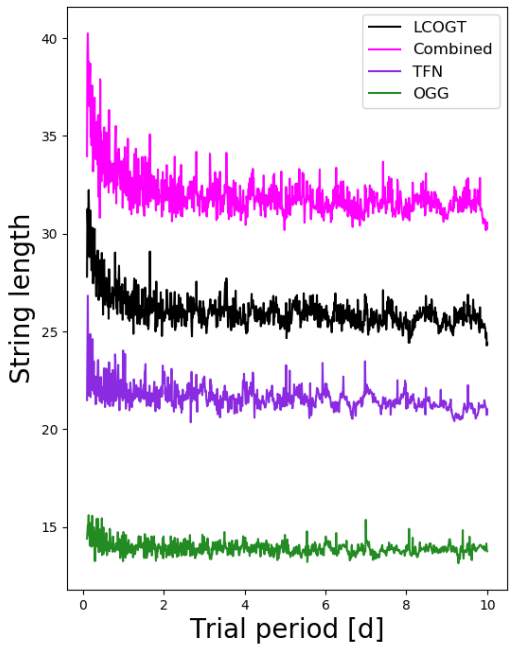

FIGURE 2.12: MJD plotted against divided counts for the majority of LCOGT R-filter observations with propagated errors. The LCOGT curve has been shifted down the y axis by 4 for clarity.

Dataset	Period [d]	Θ_{PDM}	p-value
OGG	0.47	0.823	0.022
	1.61	0.866	0.067
	1.51	0.875	0.082
	6.25	0.880	0.091
	3.22	0.888	0.110
TFN	0.13	0.740	2.12×10^{-4}
	0.40	0.807	5.57×10^{-3}
	9.09	0.820	9.32×10^{-3}
	2.56	0.847	0.023
	5.56	0.865	0.042
Combined	1.00	0.791	9.3×10^{-5}
	0.27	0.803	5.1×10^{-4}
	0.72	0.815	1.3×10^{-3}
	0.24	0.833	1.7×10^{-3}
	4.00	0.833	1.7×10^{-3}

TABLE 2.6: PDM results: trial periods with lowest Θ_{PDM} values, and corresponding p-values for R-band reductions. Where multiple low-Θ_{PDM} values cluster around a similar period, only the strongest signal is detailed.

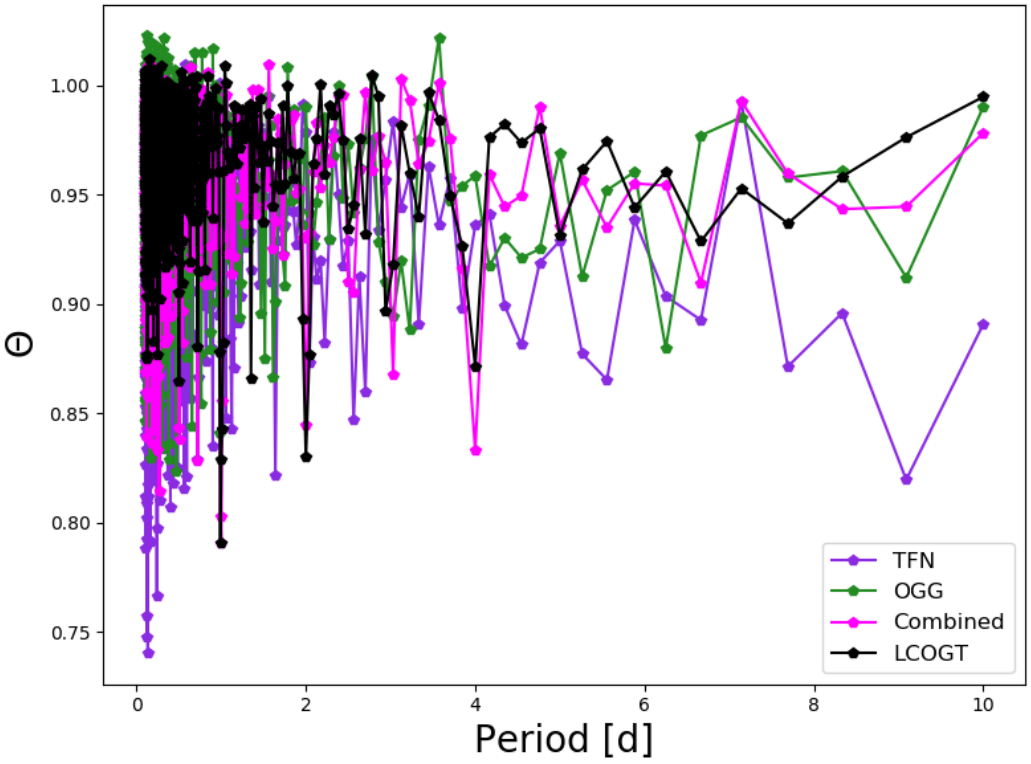

FIGURE 2.13: MJD plotted against divided counts for the majority of LCOGT R-filter observations with propagated errors.

2.5.3 Phase-folded light curves

Folding the divided counts values onto the most significant periods produces the light curves in Figure 2.14. These show OGG and TFN observations in red and green respectively, with binned points using the same colour code. Binned points for the combined dataset are shown in blue. The most convincingly sinusoidal curve is folded on a 5.6 d period, identified by the SL method for both OGG and TFN data. Barring two bins, the OGG curve folded onto a 5.27 d period is sinusoidal in character. The TFN curve folded onto an 8.44 d period also appears sinusoidal.

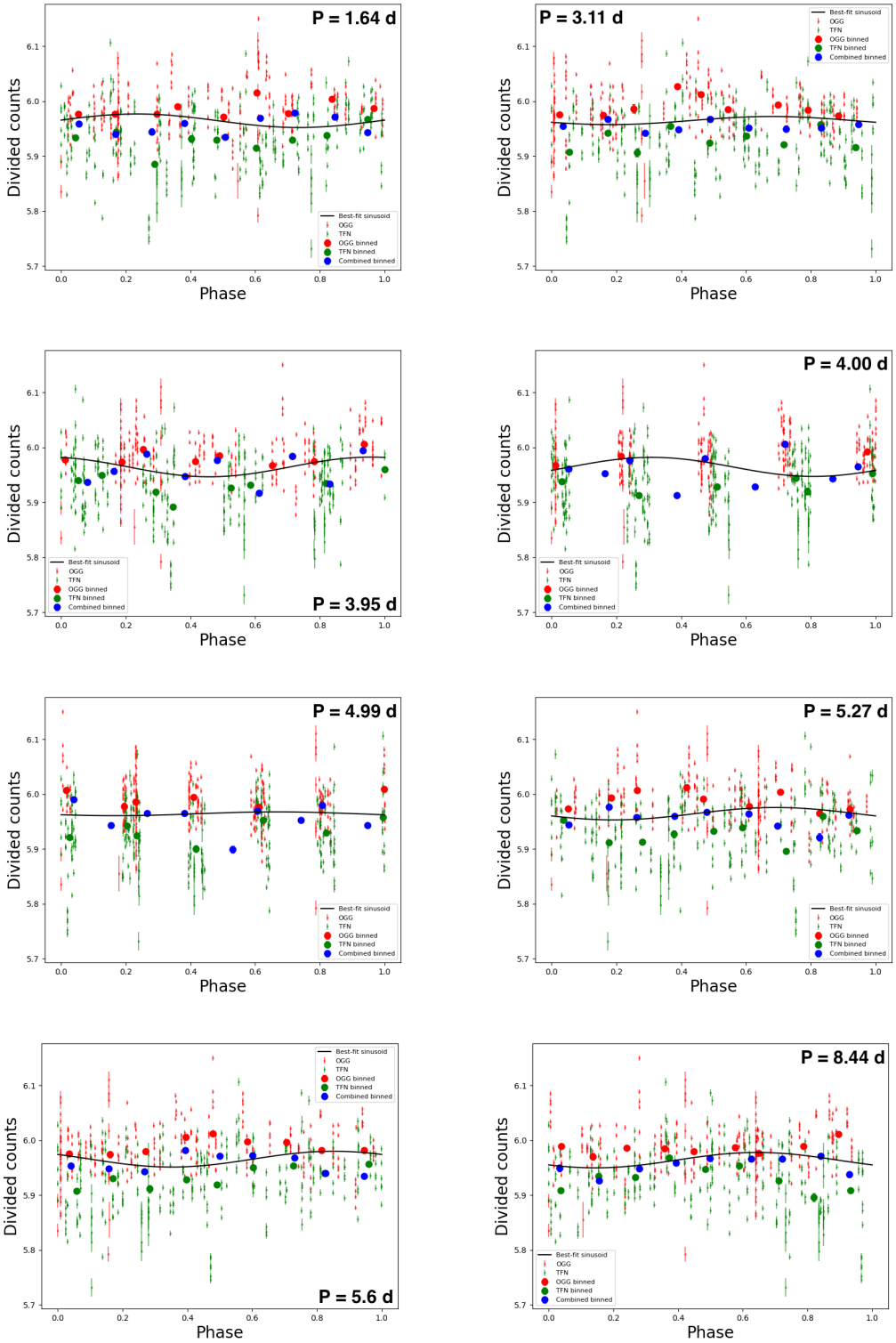

FIGURE 2.14: Phase-folded light curves for R-filter data with OGG and TFN binned separately (red and green circles) and together (blue circles).

2.6 Discussion

From the preceding analysis, the most promising signals have $\approx 5-6$ d periods. Taking $P_{\rm orb} = 5.6$ d and $R_{\rm P} = 1.2\,{\rm R}_{\odot}$, stellar inclination i may be calculated to be $i = 44^{\circ}{}^{+4}_{-13}$. This indicates HD 184960's orbit may be significantly inclined. Photometric variability may also be constrained. The phase-folded light curves presented may be compared to those of the Pleiades cluster, a selection of which are shown in Figure 2.15.

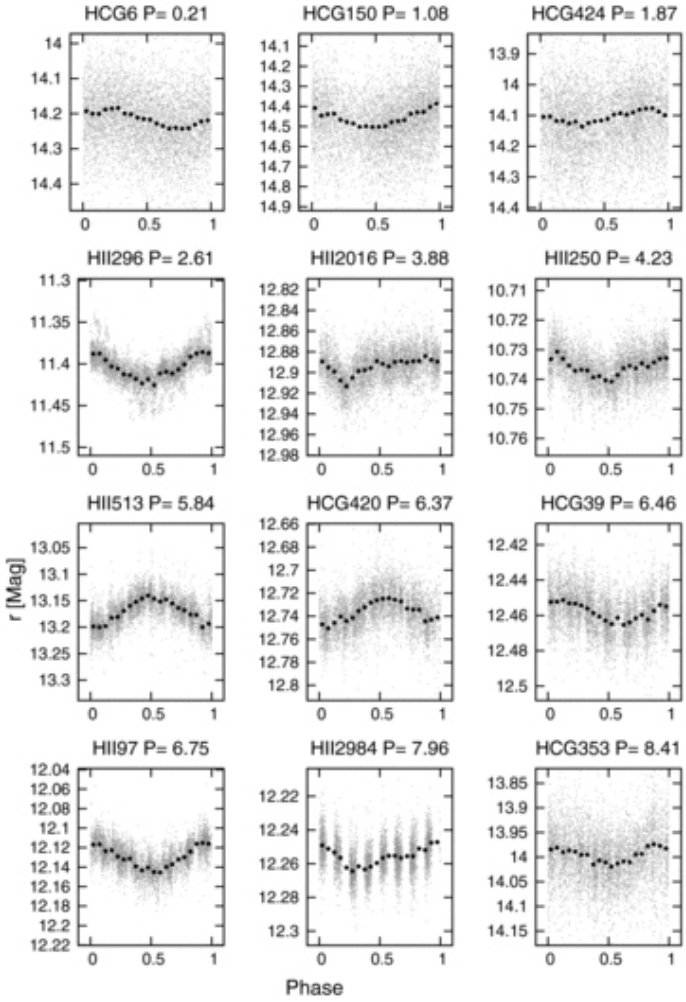

FIGURE 2.15: Phased light curves for 12 stars randomly selected from a catalogue of Pleiades stars, sorted by rotation period. The grey-scale points show all the photometric data, the dark filled circles show the phase-binned light curve. (Hartman, 2010)

Pleiades stars are significantly younger and, therefore, more active than HD 184960. More photometric variability means amplitudes should be larger and represent an upper limit for our observations. The Pleiades phase-folded curves are more sinusoidal in nature on inspection, with amplitudes that range from $0.0194 - 0.0916$ mag in the R-band (Hartman, 2010). Figure 2.16 shows that, while lower amplitude variations in Pleiades are found at all periods, they tend to cluster at bluer colours. Stars bluer than $(V - K_s)_0 < 1.1$ have distinctly lower amplitudes. HD 184960 has $(V - K) = 1.2$, so Figure 2.16 indicates a possible amplitude of $0.001 - 0.020$ mag.

To compute the detection limit of the set-up presented, I first calculated the $3\,\sigma$ variance of the binned points in the light curves for the most promising period signals, and converted these to percentages of the mean divided counts d_i. Calculating these percentages for the quotient of HD 184960 and HD 185147 B magnitudes (0.698), allows for comparison to Figure 2.16 (although note HD 185147 does not have a published R-magnitude.) The calculations are presented in Table 2.7. Amplitudes of 0.015 mag for B-filter data, 0.009 mag for R-filter data, and 0.018 mag for LCOGT extracted data may be detected with the current set-up. This constrains the photometric amplitude of HD 184960 to $\lesssim 0.018$ mag.

Filter	Dataset	Period [d]	$3\,\sigma$ [counts]	$3\,\sigma$ [%]	B amplitude [mag]
B	Ap10	5.01	0.29	2.52	0.017
B	LCOGT	5.13	0.25	2.14	0.015
R	OGG	5.27	0.075	1.26	0.009
R	TFN	5.60	0.0817	1.47	0.010
B	LCOGT	6.25	0.30	2.57	0.018

TABLE 2.7: Table detailing the detection limits for HD 184960. Filter, dataset, and period are detailed in the first three columns. The $3\,\sigma$ peak-to-peak amplitude of the divided counts is presented in counts and as a percentage. The photometric amplitude detectable with the current set-up is presented in the final column.

There are a number of possible reasons why a definitive rotation period has not been identified. HD 184960 may have intrinsically low activity and subtle photometric variability. Small starspots that induce modest photometric changes may have evolved over the time scale of observations, confounding identification of the rotation period. The star is very bright, so short exposure times were required to

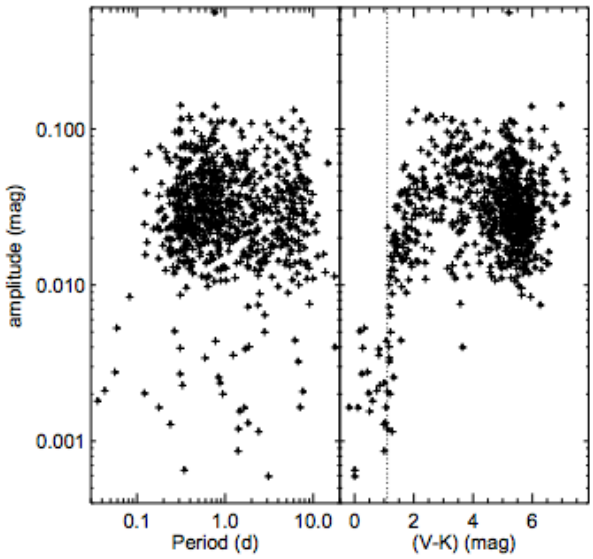

FIGURE 2.16: The amplitude, in magnitudes, of the periodic light curves against P and $(V - K_s)_0$. The vertical dotted line is at $(V - K_s)_0 = 1.1$. Stars bluer than about $(V - K_s)_0 \sim 1.1$ have lower amplitudes. The median amplitude is 0.030 mag. (Rebull et al., 2016a,b)

avoid CCD saturation. This means instrumental instabilities were not averaged out over long exposures. Due to the limited FOV, only one reference star was available. Variations in HD 185147's brightness introduced further uncertainty into our measurements. However, the late-type $V-K = 2.16$ star has no significant Ca II H & K emission ($\log(R'_{\rm HK}) = -5.21$: Strassmeier et al., 2000), so its photometric amplitude should be small. A larger FOV that captures more bright stars would be beneficial. Systematic effects between LCOGT telescopes have also introduced additional uncertainty. While it is beneficial to have a wider observation window, a single instrument is preferable for making precision measurements. High cadence observations spanning several rotation periods with a stable instrument, such as the Transiting Exoplanet Survey Satellite (TESS: Ricker et al., 2014), could help to identify the rotation period of this low amplitude star.

Update (09/2020): HD 185147 is in the TESS dataset. Photometry and a periodogram search indicates a signal at 6.33 d with FAP $= 3.2 \times 10^{-25}$, and likelihood function $\Delta \log L = 62$ where (from log-likelihood periodgram fitting) values above 15 are significant. The amplitude of the variation is low, with the ratio of the change in flux to the mean flux $\Delta F / \bar{F} = 0.0047\%$. This corresponds to a star spot of radius 0.3°.

2.7 Summary

We observed HD 184960 in B- and R-bands with 0.4 m telescopes of the Las Cumbres Observatory Global Telescope (LCOGT). Due to the brightness of the target, the telescopes were defocused. A greater degree of defocusing was used for R-band observations. Reduction and photometry were performed by LCOGT's pipeline, with photometry repeated using IRAF and AstroImajeJ software for B-band and R-band observations respectively. IRAF removed spurious artefacts from LCOGT pipeline outputs. Only one star in the FOV (HD 185147) was bright enough to act as a reference star. This was divided into the target to give 'divided counts' (d_i). No correlations with airmass were identified.

I used three algorithms to search for period signals in the photometry outputs: the Lomb Scargle (LS) periodogram, phase dispersion minimization (PDM), and the string-length (SL) method. The LS periodogram identified several high significance, low-power peaks (e.g. $P_{\rm rot} = 6.07$ d, $p = 5.36 \times 10^{-5}$, LS power$= 0.061$). The PDM and SL methods also identified possible periods (e.g $P_{\rm rot} = 4.18$, $p = 0.003$ and $P_{\rm rot} = 5.00$, $p = 0.049$ respectively). Divided counts were folded onto the most promising periods, binned, and plotted. While there was no stand-out period, several candidate signals with $P_{\mathbf{rot}} \approx 5 - 6$ d produced sinusoidal-like variability. This indicates that the star is significantly inclined ($i = 44^{\circ}{}^{+4}_{-13}$). The variance of binned points was used to identify detection limits and constrain the photometric amplitude of HD 184960 to $\lesssim 0.018$ mag. Possible improvements to the observing strategy were discussed.

Collaboration details

The original observing proposal for this work was prepared by Carole Haswell. I subsequently planned and requested B- and R-band observations. I performed B-filter photometry. R-band photometry was carried out by John Barnes. I undertook all period analysis and interpretation presented.

Chapter 3

The OU-SALT survey

I begin this chapter by advocating for a homogeneous, chromospheric activity study (Section 3.1). In Section 3.2, I describe our instrumental set-up and observations. There follows a step-by-step guide to the data reduction process adopted, which includes a discussion of specific challenges encountered (Section 3.3). Extraction of S-values and conversion to the 'end product' $\log(R'_{\rm HK})$ metric is detailed in Section 3.4. The results of the OU-SALT survey are presented in Section 3.5, and compared to published activity values in Section 3.6. In Section 3.7, I provide a summary and details of collaboration.

3.1 Survey value

Astronomers have traditionally used chromospheric activity measurements to infer fundamental properties of cool stars including age, rotation rate, convection levels and magnetic activity. With the advent of exoplanetology, activity metrics for planet-hosts may be used to probe planetary mass-loss, star-planet interactions (SPI), planet composition determinations, radio detectability and planetary magnetic fields (Doherty et al., 2019). However, the majority of exoplanet hosts lack chromospheric activity data. S values (see Section 1.3.3) are provided for only 7% of confirmed planet hosts on the NASA Exoplanet Archive[1]. Discovery papers for planets found by Kepler (Borucki et al., 2010) do not generally detail activity values, while ground-based transit surveys such as Super-WASP (Pollacco et al., 2006) and HAT-S (Bakos et al., 2013) routinely publish planet discoveries without activity values.

Where activity data is published, it is heterogeneous having been measured with a variety of spectrographs and calibrations. Large telescopes and long exposure times are required to produce adequate signal-to-noise ratio (SNR) in Ca II H & K cores but few southern hemisphere spectrographs are calibrated to generate $\log(R'_{\rm HK})$ metrics. Different calibrations to the Mount Wilson system can dominate uncertainty in high SNR $\log(R'_{\rm HK})$ measurements (Jenkins et al., 2011). Published results often lack uncertainly estimates.

To address these deficiencies, we have created the OU-SALT survey: a homogeneous study of the chromospheric activity of southern hemisphere transiting exoplanets. In this chapter, I present $\log(R'_{\rm HK})$ measurements of 104 systems from observations taken during SALT semesters 2016-1 to 2018-1. The OU-SALT database is the first homogenous activity survey of planet hosts of this scale and quality. The programme avoids uncertainties caused by systematic offsets between measurements using different instruments. We have thus produced a uniform database with significantly smaller internal errors than the prior state-of-the-art in planet-host activity work.

[1] https://exoplanetarchive.ipac.caltech.edu

I will use the OU-SALT database to test the following hypotheses:

1. Close-in planets (with semi-major axes $a < 0.1$ AU) experience mass loss;
2. Under favourable physical conditions, material stripped from close-in planets can form circumstellar enshrouding disks that absorb photospheric emission cores; and
3. Close-in planets interact with their host stars tidally and magnetically to enhance or suppress chromospheric activity.

We expect the OU-SALT database to be exploited to probe other phenomena too. Systems containing mass-losing planets are prime targets for transmission spectroscopy, while those with evidence of SPI may be observed to characterise interactions. Ca II H & K emissions may be used to predict radio brightness, which should soon be observable (See et al., 2015; Vidotto et al., 2015). Radio emissions will provide insight into rotation periods and magnetic moments, which will allow magnetic SPI to be thoroughly probed.

3.2 Observations

Staab et al. (2017) calibrated the Robert Stobie Spectrograph (RSS: Kobulnicky et al., 2003, Figure 3.1) at the 10 m South African Large Telescope (SALT: Buckley et al., 2006) to efficiently measure stellar activity. The RSS offers a choice of wavelength coverage and spectral resolution. A 0.6 arcseconds slit was selected to produce spectral resolution $R \sim 7300$ ($R > 2500$ is required for precision $\log(R'_{\rm HK})$ measurements: Jenkins et al., 2011) at 4000 Å, with wavelength coverage from 3882 Å to 4614 Å. A camera station angle of 79.75° and grating angle of 39.875° were chosen to ensure the wavelength range of interest avoided gaps between three charged couple devices (CCDs). A Copper-Argon (CuAr) arc-lamp provided calibration lines near the Ca II H & K cores. Exposures of the lamp of $\sim$ 50 s were taken immediately after science images. SNR > 15 in Ca II H & K cores was achieved with $\sim$ 400 s exposures of $V \lesssim 13$ targets, even in poor observing conditions.

I prepared and submitted observing proposals via the Principal Investigator Proposal Tool (PIPT)[2] for SALT semesters 2016-2 to 2019-2, specifying the settings described above **for each target**. Observations were made by resident SALT astronomers. Bright targets ($V \lesssim 12$) were observed with 400 s exposure times. Fainter targets ($V > 12$) were observed for $1200 - 1600$ s blocks, split into $3 - 4$ separate exposures to minimise the risk of cosmic ray strikes in Ca II H & K cores. Our SNR requirements were generally met where the full-width-half-maximum (FWHM) of the seeing disc was 2.5 arcsecs or less.

3.3 Data reduction

The SALT pipeline (Crawford et al., 2010) corrected science frames for CCD bias and gain. I have taken the following further steps to reduce SALT data:

[2] https://astronomers.salt.ac.za/software/pipt/

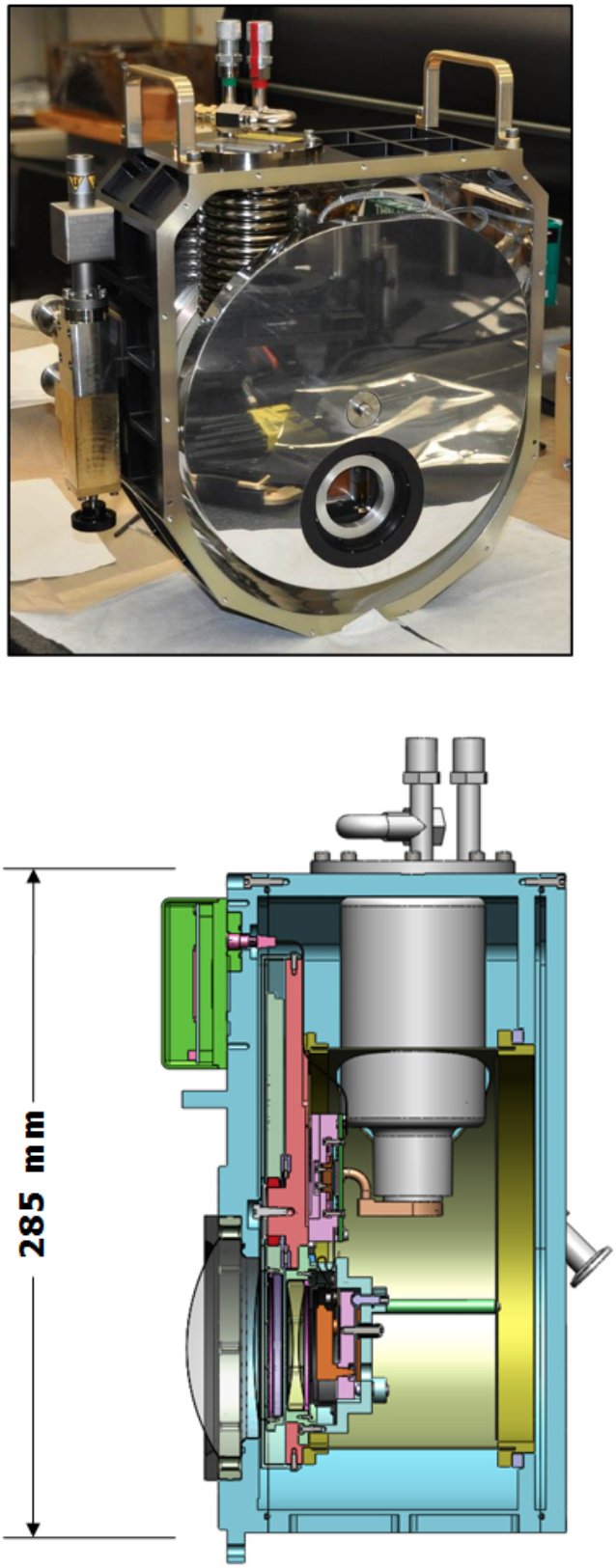

FIGURE 3.1: Photograph and schematic of the Robert Stobie Spectrograph (Credit: John Hopkins University.)

3.3.1 Image trimming

Science, arc and flat frames were trimmed in the dispersion (x) and spatial (y) axes to ensure all images had the same dimensions. The target is located centrally in the y-direction, meaning the top and bottom of images were largely redundant provided sufficient regions remained for background subtraction.

3.3.2 Wavelength calibration

Longslit spectroscopy images are geometrically distorted. This is illustrated by Figure 3.2, which shows CuAr lines that are curved relative to the CCD columns. It was necessary to correct science frames using a 2D wavelength solution. Correction ensured that the background fit was subtracted from the matching wavelength range on the stellar trace. Wavelength calibration involved a number of steps using the *twospec.longslit* package in the Image Reduction and Analysis Facility (IRAF: Tody 1986).

In order to calibrate science spectra, I manually identified and labelled the wavelengths of six separate strong spectral lines (at 3928.623 Å, 4131.724 Å, 4158.592 Å, 4348.064 Å, 4545.05 Å and 4609.567 Å) across the width of the CuAr spectrum corresponding to the observation. These lines are published by SALT[3]. This provided sufficient information to automatically label the remaining 37 lines specified in a separate file, which I chose because they are in the correct wavelength range and are reasonably straight forward to identify. I fitted a one dimensional wavelength solution at a single position along the y axis using the *identify* task. A third order cubic spline generally provided solutions with low RMS and no systematic effects in the residuals. Where the fit for a particular line had RMS < 0.05 Å, it was removed from the solution.

I employed the *reidentify* task to re-fit the solution in 10 pixel increments along the spatial axis of the CuAr frame. This effectively tracked line curvature. I used this output and the *fitcoords* task to compute x and y coordinates of the 'surface-defining'

[3] http://pysalt.salt.ac.za/lineatlas/CuAr.txt

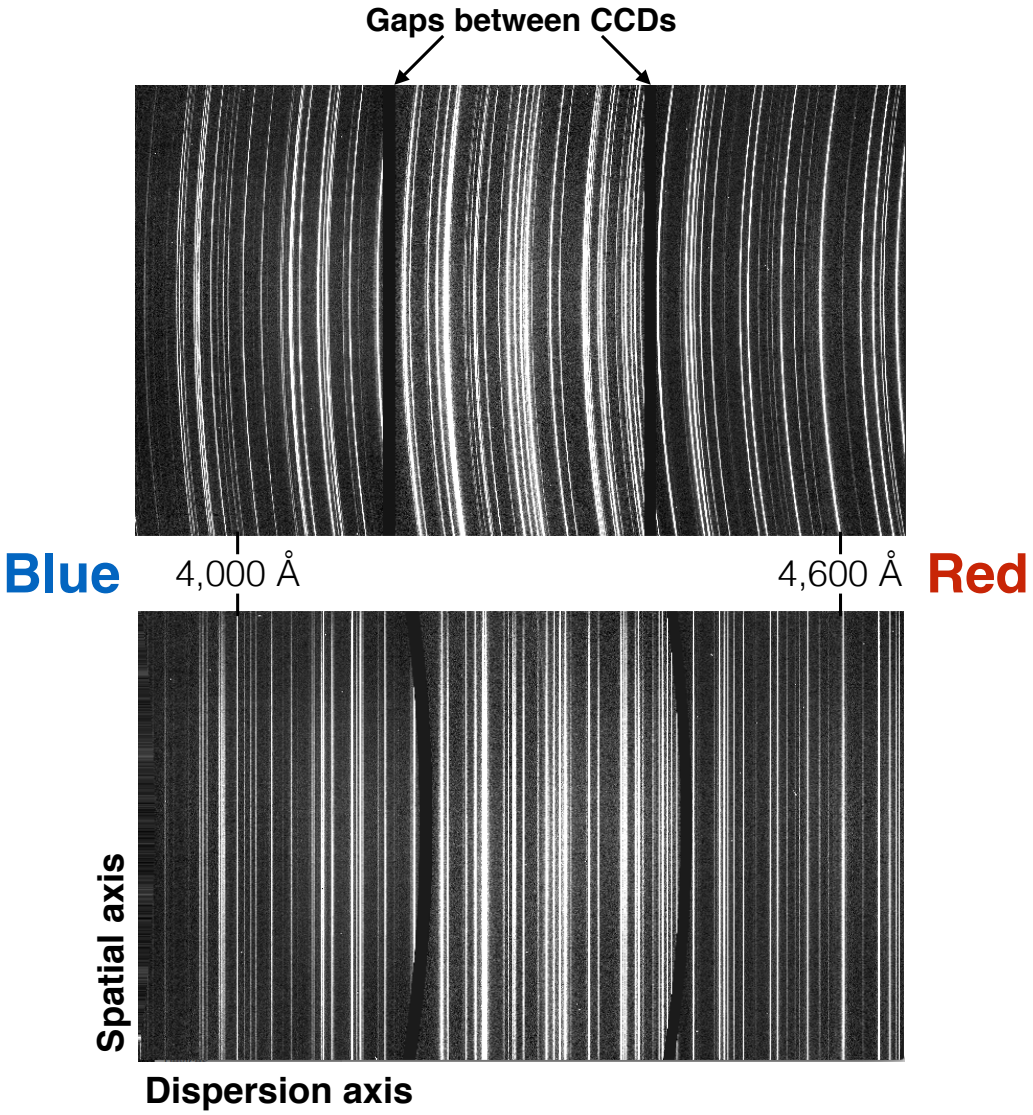

FIGURE 3.2: Geometrically distorted RSS CuAr arc lamp image (top), which I corrected using IRAF's *twospec.longslit* package (bottom). An approximate wavelength scale is indicated, and the gaps between CCDs are annotated.

wavelength on the CCD. These positions were used to geometrically correct exposures with the *transform* task. The x and y axes are linear functions of wavelength and position respectively.

3.3.3 Flat-fielding

Flat-field frames were used to remove pixel-to-pixel variations across the CCD. One set of five flats matching our spectral and detector settings were provided each semester. SALT astronomers reported that the flat-field is stable on time-scales of weeks. Flats were median combined to produce a master flat frame, which was divided into the science frames. Flat fielding was particularly important to correct for a column of lazy pixels on the CCD that returned lower-than-expected counts. The column, located around $\sim$ 3994Å, intersected with the R continuum passband required to calculate the S_{RSS}-value. This crucial correction was specifically checked in each science frame, with the column visually apparent if not corrected.

3.3.4 Cosmic ray removal

I used the *L.A.Cosmic* algorithm (Van Dokkum, 2001) to remove cosmic ray artefacts from science frames. The algorithm employs Laplacian edge detection: a mathematical method to identify points where there is a sharp change in counts, i.e. there is a discontinuity. The algorithm detects cosmic rays of arbitrary shapes and sizes by the sharpness of their edges. Following flat-fielding, I specified how many times the algorithm should run (generally 2-7), depending on science frame exposure time. A significant number of artefacts were found in longer exposures. It was necessary to fine-tune the algorithm so that spectral features were not mistakenly identified and removed—particularly strong emission cores. The detection limit for cosmic rays was set to 10 σ, the fractional detection limit for neighbouring pixels to 0.5, and the contrast limit between the cosmic ray and underlying object to 4.0. Cosmic ray artefacts in the stellar spectrum that might have survived this cleaning process were accounted for during the spectral extraction step detailed in the following section.

3.3.5 Background subtraction and spectral extraction

Accurate background subtraction is crucial to producing reliable $\log(R'_{\rm HK})$ values as systematic effects can depress activity measurements. Background over-subtraction of 5% for a star near the basal limit leads to a 50% reduction in the measured S-value (Fossati et al., 2017b). The solar spectrum, via scattered moonlight, dominates the background.

I employed the *apall* function to carry out background subtraction and extract stellar spectra. Once the centre of the spectrum was identified along the spatial axis, and a suitable aperture width specified, I manually selected large regions of background either side of the stellar profile. I then chose a function to fit the background profile and interpolate between the background windows and the aperture. Background intensity distribution varied across science frames, so different order fits were required: second and third order Chebyshev functions (Apostol, 1976) generally proved the most effective fit. I checked the fit at four points along the dispersion axes (at 300, 1000, 2000 and 2500 pixels) to ensure goodness of fit across the CCD, particularly where scattered light was present in the science image. The fit within the region of the Ca II H & K was given priority over the fit elsewhere on the spectrum.

The *optimal* extraction option within *apall* was used to carry out variance-weighted extraction of stellar spectra, with most weight given to data points with the smallest uncertainty (Horne, 1986). Spectra are traced across the dispersion axis of CCD frames. For each wavelength sampling position in the spatial direction, the counts within the aperture were added and the background level subtracted. Pixels within the extraction and background fit windows were rejected if affected by cosmic ray hits, providing a further round of cosmic ray scanning in addition to that carried out by *L.A.Cosmic*. Photon and detector readout noise were propagated to provide uncertainties on the extracted spectra. The *apall* task outputs text files of optimised science and error spectra (i.e. counts and CCD y-coordinates), along with normalised versions, sky spectra and processed CCD images.

3.3.6 Specific challenges

Various challenges arose during reduction of OU-SALT data:

1. Several science frames were not usable due to RSS spectral focus issues. These observations were repeated.

2. Unexplained artefacts in the observation of WASP-121 unfortunately rendered the science frame for this extremely interesting target unusable. The artefacts appeared to be pixel malfunctions in key spectral regions. The observations have been re-scheduled.

3. Each image is composed of three rectangular regions separated by two vertical gaps, which result from SALT's CCD array (see Figure 3.3). While these gaps are located away from key areas of interest, the images were nevertheless interpolated at the start of the reduction process.

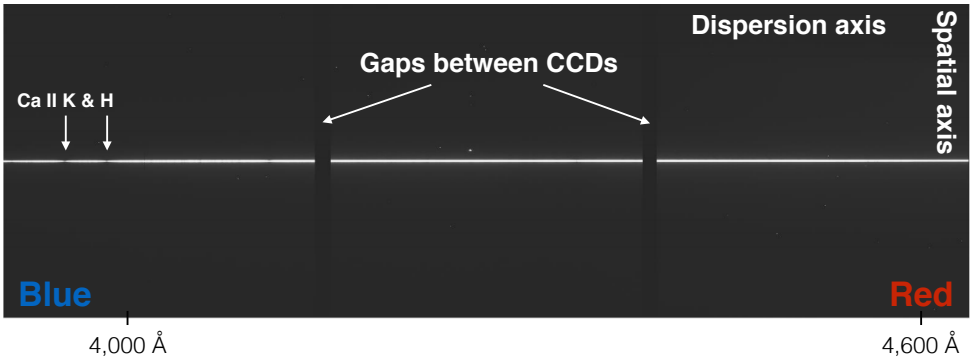

FIGURE 3.3: Raw image of calibrator star HD 182101 on the 0.6 arcsecond slit of the RSS. The stellar trace is the bright streak spanning the three CCDs. An approximate wavelength scale is indicated. The Ca II H & K lines are apparent even in this raw image.

4. Pixel dimensions of frames changed between 2016-1 and 2016-2 semesters meaning flats and science frames were different sizes. IRAF's *imcopy* function was used to trim images so that flat-fielding remained effective.

5. Arc lamp exposures are too faint at the top and bottom to provide reliable wavelength solutions (see Figure 3.4). All exposures were trimmed for more reliable 2D solutions.

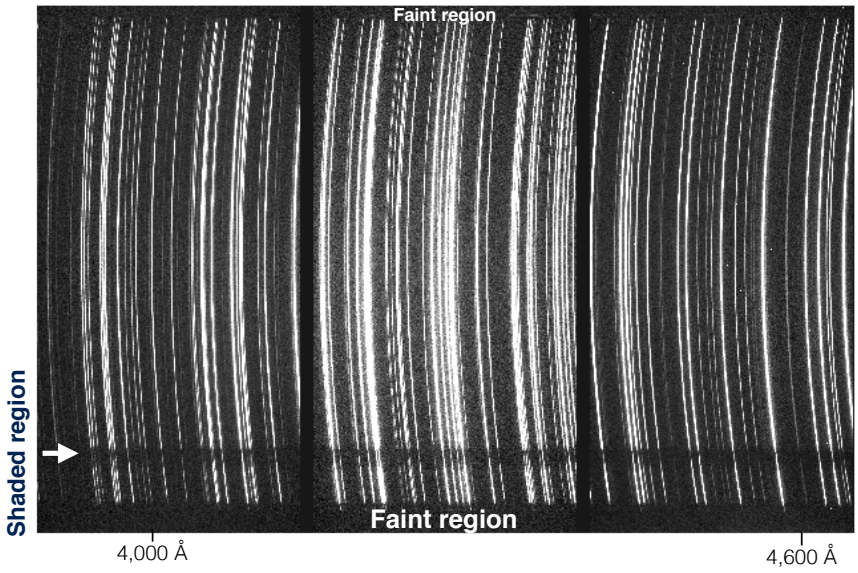

FIGURE 3.4: CuAr arc lamp for HD 3167, showing faint regions at top and bottom of image, and a horizontal shadow in the bottom quarter. An approximate wavelength scale is indicated. Dispersion and spatial axes as per Figure 3.2.

6. Multiple arc frames have horizontal shadows where dust has settled around the slit (Figure 3.4). These areas were trimmed provided there was sufficient background remaining for subtraction.

7. The $\log(R'_{\rm HK})$ metric is highly sensitive to flux in the Ca II H & K cores (Staab et al., 2017). Direct CR hits to cores renders an image unreliable for producing $\log(R'_{\rm HK})$. CR hits in the passbands may be reliably removed without major impact on measured $\log(R'_{\rm HK})$.

8. Spectra may be extracted for multiple stars falling on the slit with careful sky background subtraction. For example, activity metrics for WASP-94 A & B (Neveu-VanMalle et al., 2014) were produced from the same image.

3.4 S-value extraction and conversion to $\log(R'_{\rm HK})$

A detailed technical description of the process for extracting S values from OU-SALT observations is provided in Staab et al. (2017). In summary, after spectra are shifted into the stellar rest frame using PyAstronomy's *crosscorrRV* routine[4], flux values are extracted from the Ca II H & K cores, and R & V continuum windows (Figure 3.5).

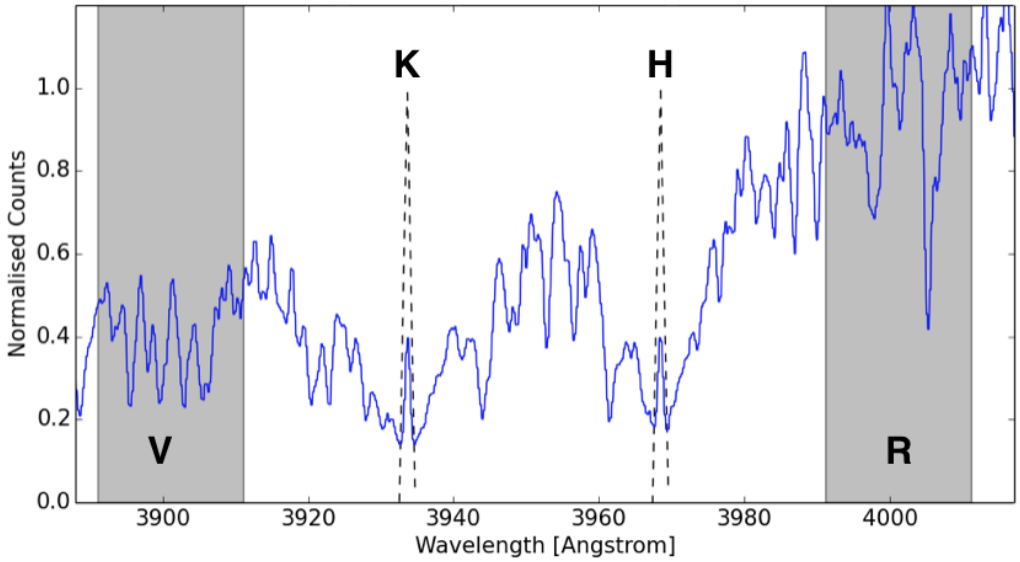

FIGURE 3.5: The extracted spectrum of WASP-52, which contains spectral features and measurement noise. I used triangular core bandpasses, shown with dashed lines, to extract flux values from the Ca II H & K cores. The R and K continuum windows are shaded. I normalised counts to the mean of the R continuum bandpass.

The instrumental S-index ($S_{\rm RSS}$) is the ratio of flux in the core bandpasses to flux in the continuum windows.

$$S_{\rm RSS} = \frac{H + K}{R + V} \,. \tag{3.1}$$

Flux calibration is unnecessary as bandpass placement renders $S_{\rm RSS}$ insensitive to the local spectral slope (Gray et al., 2003; Fossati et al., 2017b).

[4] https://github.com/sczesla/PyAstronomy

The following relation from Staab et al. (2017) was used to calibrate to the Mount Wilson system:

$$S_{\rm MW} = \frac{S_{\rm RSS} - (0.075 \pm 0.005)}{(0.60 \pm 0.02)}. \tag{3.2}$$

Noyes et al. (1984) provides a mathematical description for converting $S_{\rm MW}$ to $R'_{\rm HK}$, being the normalised chromospheric contribution contribution to $R_{\rm HK}$. This requires stellar $B-V$ which, for consistency, I calculated using stellar parameters reported in the Transiting ExoPlanet CATalogue (TEPCat: Southworth, 2011) and the following expression:

$$\begin{aligned}(B-V) = {} & t_0 + t_1 \log T_{\rm eff} + t_2 (\log T_{\rm eff})^2 \\ & + t_3 (\log T_{\rm eff})^3 + f_1 [{\rm Fe/H}] + f_2 [{\rm Fe/H}]^2 \\ & + d_1 [{\rm Fe/H}] \log T_{\rm eff} + g_1 \log g \\ & + e_1 \log g \log T_{\rm eff},\end{aligned} \tag{3.3}$$

where coefficients are listed in Table 4 of Sekiguchi & Fukugita (2000). The colour-dependent conversion factor $C_{\rm cf}(B-V)$ from Middelkoop (1982):

$$\log C_{\rm cf} = 1.13(B-V)^3 - 3.91(B-V)^2 + 2.84(B-V) - 0.47, \tag{3.4}$$

converts the S index into $R_{\rm HK}$:

$$R_{\rm HK} = 1.34 \times 10^{-4} C_{\rm cf} \times S_{\rm MW}. \tag{3.5}$$

The colour-dependent photospheric contribution to the cores ($R_{\rm phot}$) is calculated:

$$\log R_{\rm phot} = -4.898 + 1.918(B-V)^2 - 2.893(B-V)^3. \tag{3.6}$$

This is subtracted in order to correct for the fraction of the total emission that is photospheric rather than chromospheric (Noyes et al., 1984). The relation defining $R'_{\rm HK}$ is thus:

$$R'_{\rm HK} = R_{\rm HK} - R_{\rm phot} \tag{3.7}$$

The $\log(R'_{\rm HK})$ metric is dimensionless, normalised by total bolometric flux of the star, and not dependent on stellar colour. It may be used to compare the chromospheric activity of F, G and K stars. Uncertainty in $R'_{\rm HK}$ is discussed in the following section.

3.5 Homogeneous activity survey

The OU-SALT homogeneous activity database for semesters 2013-2 to 2018-2 is presented in Table 3.1, with 141 observations of 104 targets. Consecutive measurements of targets within a single observing block are median combined and presented as a single observation. The final two columns show activity and gyrochronological age estimates for each host (see Section 4.3). Figure 3.6 shows $B-V$ plotted against $\log(R'_{\rm HK})$ for the each target's median activity value. There are no "activity dropouts", being outliers with line core flux lower than the photospheric flux contribution ($R_{\rm phot}$) that lead to S-values $\lesssim 0.1$ and negative R'_{HK} values (Staab et al., 2017; Staab, 2018).

3.5.1 Uncertainties

Our $\log(R'_{\rm HK})$ results are presented in Table 3.1 with both internal and external uncertainties. Published uncertainties in effective temperature ($T_{\rm eff}$), metallicity [Fe/H] and stellar surface gravity ($\log g$) are propagated into $B-V$. Uncertainties in $B-V$ are propagated into $R'_{\rm HK}$. The $S_{\rm RSS}$ values are presented with photon noise uncertainties, which are *internal* to the OU-SALT survey. Where photon noise uncertainty (and not calibration-related uncertainty) is propagated into $\log(R'_{\rm HK})$, the black error bars in Figure 3.6 result. These error bars range between $0.09 - 2.51$ % of $\log(R'_{\rm HK})$, with a median of 0.59 %. Photon noise uncertainties increase for cooler targets.

S_{MW} values are presented with calibration uncertainty as detailed in Equation 3.2. Where this calibration error (along with photon noise uncertainty) is propagated into $\log(R'_{HK})$, the red error bars in Figure 3.6 result. Uncertainty in the calibration relation dominates for hotter, low activity stars. Percentage errors are $0.5 - 9.2$ % of $\log(R'_{HK})$, with a mean of 1.97 %. **Calibration errors are therefore $\sim$ 3.3 times larger than internal uncertainties.** It is appropriate to use calibration errors where we wish to compare our results to other surveys.

Uncertainties for $\log(R'_{HK})$ are asymmetric, most noticeably for the lowest activity targets, where emission flux approaches zero. For inactive stars, variation in $\log(R'_{HK})$ results in an RMS of ~ 0.015 dex (Gomes Da Silva et al., 2014). Typical astrophysical variability is therefore largely negligible for $\log(R'_{HK})$ (Staab et al., 2017).

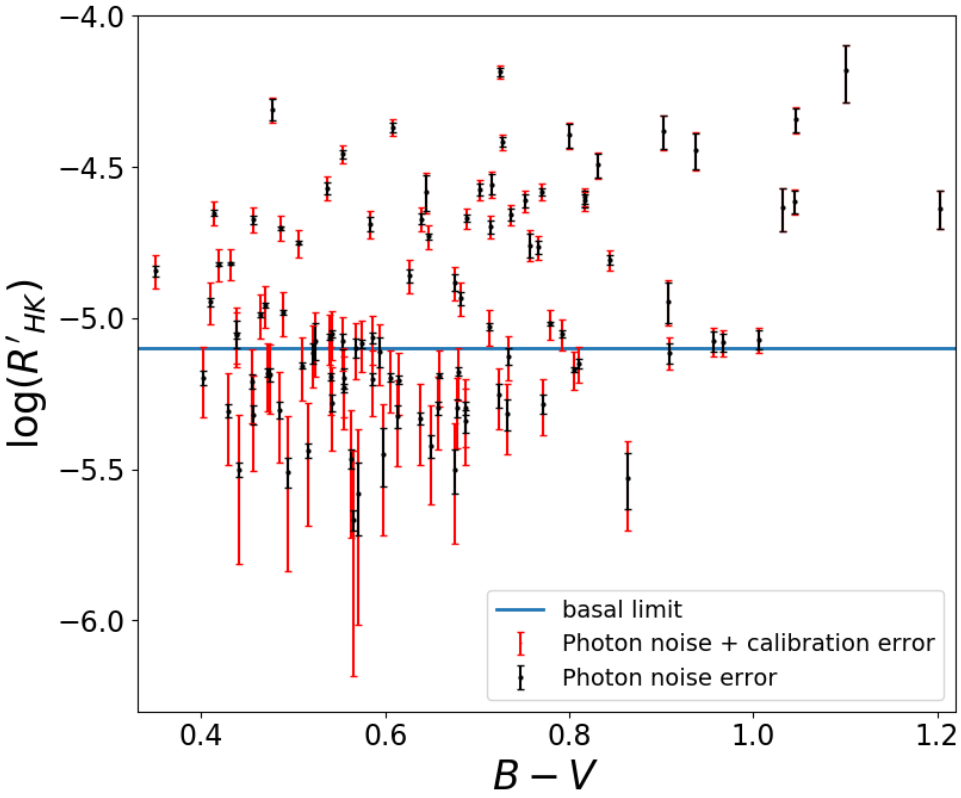

FIGURE 3.6: Colour-activity plot of median OU-SALT measurements for each target. Black error bars show internal photon noise uncertainty while red error bars show both photon noise and calibration uncertainty.

TABLE 3.1: OU-SALT measurements and derived quantities: Modified Julian Date (MJD), planetary orbital phase (ϕ), RSS S value ($S_{\rm RSS}$), Mount Wilson S value ($S_{\rm MW}$), colour index ($B-V$), $\log(R'_{\rm HK})$ with (external) calibration errors and (internal) instrumental errors, activity age, and gyro age. Median $S_{\rm RSS}$, $S_{\rm MW}$, and $\log(R'_{\rm HK})$ values are provided where consecutive observations of a target were made within a single observing block.

Target[a]	MJD −2450 K	Phase [ϕ]	$S_{\rm RSS}$	$S_{\rm MW}$	$B-V$[b]	$\log(R'_{\rm HK})$	Act Age [c] [Gyr]	Gyro age[d] [Gyr]
CoRoT-7	8419.569	0.260	0.267 ± 0.002	0.320 ± 0.014	0.82 ± 0.02	$-4.61^{+0.03}_{-0.03}\ ^{+0.021}_{-0.022}$	$1.22^{+0.25}_{-0.20}$	$1.75^{+0.55}_{-0.55}$
CoRoT-11	7563.414	0.453	0.300 ± 0.013	0.375 ± 0.026	0.48 ± 0.02	$-4.31^{+0.04}_{-0.04}\ ^{+0.033}_{-0.035}$	$0.15^{+0.06}_{-0.04}$	$2.00^{+1.00}_{-1.00}$
CoRoT-22	8238.638	0.127	0.146 ± 0.007	0.119 ± 0.015	0.60 ± 0.04	$-5.39^{+0.20}_{-0.37}\ ^{+0.158}_{-0.252}$	> 11.0	$3.30^{+2.00}_{-2.00}$
CoRoT-22	8355.263	0.081	0.144 ± 0.003	0.116 ± 0.010	0.60 ± 0.04	$-5.45^{+0.17}_{-0.27}\ ^{+0.085}_{-0.106}$	> 11.0	$3.30^{+2.00}_{-2.00}$
CoRoT-28	7962.336	0.632	0.130 ± 0.003	0.091 ± 0.010	0.86 ± 0.04	$-5.53^{+0.12}_{-0.17}\ ^{+0.083}_{-0.102}$	> 11.0	$12.0^{+1.50}_{-1.50}$
EPIC 219388192	8366.39	0.733	0.296 ± 0.004	0.368 ± 0.016	0.61 ± 0.02	$-4.37^{+0.03}_{-0.03}\ ^{+0.014}_{-0.015}$	$0.24^{+0.06}_{-0.04}$	$3.90^{+1.90}_{-1.90}$
HAT-P-27	7952.273	0.077	0.272 ± 0.002	0.327 ± 0.014	0.82 ± 0.02	$-4.60^{+0.03}_{-0.03}\ ^{+0.057}_{-0.066}$	$1.14^{+0.23}_{-0.19}$	$4.40^{+3.80}_{-2.60}$
HATS-2	7587.218	0.666	0.322 ± 0.006	0.411 ± 0.019	0.83 ± 0.04	$-4.49^{+0.04}_{-0.05}\ ^{+0.020}_{-0.021}$	$0.58^{+0.21}_{-0.15}$	$9.70^{+2.90}_{-2.90}$
HATS-3	8239.601	0.295	0.153 ± 0.001	0.131 ± 0.010	0.46 ± 0.02	$-5.21^{+0.11}_{-0.14}\ ^{+0.038}_{-0.042}$	$10.12^{+1.05}_{-1.61}$	$3.20^{+0.60}_{-0.40}$
HATS-10	7986.444	0.404	0.149 ± 0.002	0.123 ± 0.010	0.61 ± 0.04	$-5.32^{+0.12}_{-0.17}\ ^{+0.024}_{-0.026}$	$11.10^{+1.03}_{-1.03}$	$3.30^{+1.70}_{-1.70}$
HATS-13	7952.374	0.576	0.220 ± 0.003	0.242 ± 0.013	0.71 ± 0.02	$-4.70^{+0.04}_{-0.04}\ ^{+0.036}_{-0.039}$	$1.96^{+0.43}_{-0.35}$	$2.50^{+1.70}_{-1.70}$
HATS-18	8303.267	0.196	0.257 ± 0.004	0.304 ± 0.014	0.72 ± 0.04	$-4.56^{+0.04}_{-0.04}\ ^{+0.020}_{-0.021}$	$0.88^{+0.27}_{-0.20}$	$4.20^{+2.20}_{-2.20}$
HATS-21	7952.519	0.254	0.148 ± 0.002	0.121 ± 0.010	0.69 ± 0.02	$-5.34^{+0.11}_{-0.15}\ ^{+0.032}_{-0.034}$	$11.15^{+0.77}_{-0.77}$	$2.30^{+1.70}_{-1.70}$
HATS-27	7936.337	0.599	0.204 ± 0.002	0.216 ± 0.012	0.46 ± 0.02	$-4.67^{+0.04}_{-0.04}\ ^{+0.036}_{-0.039}$	$1.73^{+0.45}_{-0.34}$	$2.30^{+0.22}_{-0.22}$
HATS-29	7979.456	0.715	0.150 ± 0.002	0.125 ± 0.010	0.68 ± 0.04	$-5.30^{+0.10}_{-0.13}\ ^{+0.014}_{-0.015}$	> 11.0	$5.50^{+2.60}_{-1.70}$
HATS-30	7918.626	0.025	0.164 ± 0.001	0.147 ± 0.010	0.59 ± 0.02	$-5.06^{+0.07}_{-0.09}\ ^{+0.030}_{-0.032}$	$7.80^{+1.53}_{-1.38}$	$2.30^{+1.20}_{-1.20}$
HATS-36	8304.571	0.138	0.257 ± 0.004	0.304 ± 0.014	0.55 ± 0.02	$-4.46^{+0.03}_{-0.03}\ ^{+0.017}_{-0.018}$	$0.45^{+0.11}_{-0.09}$	$3.40^{+1.90}_{-1.40}$
HD 73256	7766.368		0.292 ± 0.0004	0.362 ± 0.015	0.73 ± 0.02	$-4.46^{+0.03}_{-0.03}\ ^{+0.014}_{-0.015}$	$0.48^{+0.10}_{-0.08}$	
HD 73256	7772.59		0.312 ± 0.0003	0.394 ± 0.016	0.73 ± 0.02	$-4.42^{+0.03}_{-0.03}\ ^{+0.009}_{-0.009}$	$0.34^{+0.07}_{-0.06}$	
HD 73256	7792.301		0.298 ± 0.0003	0.372 ± 0.015	0.73 ± 0.02	$-4.45^{+0.03}_{-0.03}\ ^{+0.014}_{-0.015}$	$0.43^{+0.09}_{-0.07}$	
HD 73256	7848.381		0.317 ± 0.0004	0.404 ± 0.016	0.73 ± 0.02	$-4.40^{+0.03}_{-0.03}\ ^{+0.015}_{-0.015}$	$0.31^{+0.07}_{-0.05}$	
K2-2(1)	7683.312	0.020	0.217 ± 0.001	0.236 ± 0.012	0.84 ± 0.02	$-4.81^{+0.03}_{-0.03}\ ^{+0.015}_{-0.015}$	$3.29^{+0.51}_{-0.44}$	1.30
K2-19	7561.216	0.411	0.249 ± 0.004	0.290 ± 0.015	0.75 ± 0.02	$-4.61^{+0.03}_{-0.04}\ ^{+0.015}_{-0.015}$	$1.22^{+0.28}_{-0.22}$	$1.89^{+0.26}_{-0.26}$
K2-31	8237.634	0.518	0.263 ± 0.002	0.314 ± 0.014	0.77 ± 0.01	$-4.58^{+0.03}_{-0.03}\ ^{+0.017}_{-0.018}$	$1.02^{+0.19}_{-0.15}$	$1.43^{+0.17}_{-0.17}$
K2-32	7925.498	0.940	0.169 ± 0.001	0.157 ± 0.010	0.79 ± 0.02	$-5.05^{+0.05}_{-0.05}\ ^{+0.022}_{-0.023}$	$7.57^{+0.99}_{-0.91}$	$7.90^{+4.50}_{-4.50}$
K2-32	7952.427	0.935	0.166 ± 0.002	0.151 ± 0.010	0.79 ± 0.02	$-5.08^{+0.05}_{-0.06}\ ^{+0.012}_{-0.012}$	$8.10^{+1.00}_{-0.95}$	$7.90^{+4.50}_{-4.50}$
K2-39	7696.360	0.368	0.190 ± 0.002	0.191 ± 0.011	1.01 ± 0.03	$-5.07^{+0.04}_{-0.04}\ ^{+0.013}_{-0.013}$	$7.91^{+0.77}_{-0.73}$	$3.09^{+0.92}_{-0.70}$
K2-96(2)	8008.479	0.932	0.157 ± 0.001	0.137 ± 0.010	0.80 ± 0.02	$-5.17^{+0.06}_{-0.07}\ ^{+0.014}_{-0.014}$	$9.58^{+0.86}_{-0.93}$	$7.80^{+4.30}_{-4.30}$
K2-106	8093.283	0.509	0.152 ± 0.002	0.128 ± 0.010	0.72 ± 0.02	$-5.25^{+0.09}_{-0.11}\ ^{+0.030}_{-0.032}$	$10.62^{+0.57}_{-1.11}$	$1.46^{+0.42}_{-0.42}$
K2-135(3)	8298.605	0.017	0.650 ± 0.004	0.959 ± 0.034	1.20 ± 0.04	$-4.64^{+0.06}_{-0.07}\ ^{+0.037}_{-0.041}$	$1.41^{+0.63}_{-0.41}$	$0.49^{+0.13}_{-0.13}$
KELT-10	7860.550	0.537	0.155 ± 0.001	0.133 ± 0.010	0.59 ± 0.02	$-5.20^{+0.10}_{-0.12}\ ^{+0.019}_{-0.020}$	$10.01^{+1.08}_{-1.48}$	$4.50^{+0.70}_{-0.70}$
KELT-11	7792.418	0.241	0.170 ± 0.0004	0.158 ± 0.010	0.78 ± 0.01	$-5.04^{+0.05}_{-0.05}\ ^{+0.005}_{-0.005}$	$7.39^{+0.99}_{-0.90}$	$3.53^{+0.01}_{-0.01}$
KELT-11	7799.419	0.719	0.170 ± 0.0003	0.158 ± 0.010	0.78 ± 0.01	$-5.04^{+0.05}_{-0.05}\ ^{+0.004}_{-0.004}$	$7.43^{+0.99}_{-0.91}$	$3.53^{+0.01}_{-0.01}$
KELT-11	7800.399	0.926	0.173 ± 0.001	0.163 ± 0.010	0.78 ± 0.01	$-5.02^{+0.05}_{-0.05}\ ^{+0.006}_{-0.006}$	$6.92^{+0.96}_{-0.86}$	$3.53^{+0.01}_{-0.01}$
KELT-11	7802.390	0.346	0.174 ± 0.0004	0.165 ± 0.010	0.78 ± 0.01	$-5.01^{+0.05}_{-0.05}\ ^{+0.005}_{-0.005}$	$6.75^{+0.94}_{-0.84}$	$3.53^{+0.01}_{-0.01}$
KELT-11	7803.404	0.560	0.173 ± 0.0004	0.163 ± 0.010	0.78 ± 0.01	$-5.02^{+0.05}_{-0.05}\ ^{+0.005}_{-0.005}$	$6.94^{+0.96}_{-0.87}$	$3.53^{+0.01}_{-0.01}$
KELT-11	7814.371	0.876	0.175 ± 0.001	0.167 ± 0.010	0.78 ± 0.01	$-5.00^{+0.04}_{-0.05}\ ^{+0.006}_{-0.006}$	$6.57^{+0.93}_{-0.83}$	$3.53^{+0.01}_{-0.01}$
KELT-11	7819.334	0.924	0.174 ± 0.001	0.164 ± 0.010	0.78 ± 0.01	$-5.01^{+0.05}_{-0.05}\ ^{+0.006}_{-0.006}$	$6.82^{+0.95}_{-0.85}$	$3.53^{+0.01}_{-0.01}$

Target[a]	MJD	Phase	$S_{\rm RSS}$	$S_{\rm MW}$	$B-V$[b]	$\log(R'_{\rm HK})$	Act Age [c]	Gyro age[d]
	-2450 K	[ϕ]					[Gyr]	[Gyr]
KELT-11	7845.272	0.401	0.167 ± 0.0003	0.154 ± 0.010	0.78 ± 0.01	$-5.06^{+0.05}_{-0.06}{}^{+0.004}_{-0.004}$	$7.80^{+1.00}_{-0.93}$	$3.53^{+0.01}_{-0.01}$
KELT-11	7848.461	0.074	0.165 ± 0.0004	0.150 ± 0.010	0.78 ± 0.01	$-5.09^{+0.05}_{-0.06}{}^{+0.003}_{-0.003}$	$8.20^{+1.00}_{-0.95}$	$3.53^{+0.01}_{-0.01}$
KELT-11	7886.324	0.068	0.172 ± 0.0004	0.162 ± 0.010	0.78 ± 0.01	$-5.02^{+0.05}_{-0.05}{}^{+0.005}_{-0.005}$	$7.00^{+0.97}_{-0.87}$	$3.53^{+0.01}_{-0.01}$
KELT-15	7898.204	0.016	0.140 ± 0.001	0.107 ± 0.009	0.56 ± 0.02	$-5.67^{+0.23}_{-0.52}{}^{+0.032}_{-0.035}$	> 11.0	$4.60^{+0.50}_{-0.40}$
Qatar-2(4)	7546.229	0.393	0.705 ± 0.008	1.049 ± 0.038	1.05 ± 0.02	$-4.34^{+0.04}_{-0.04}{}^{+0.037}_{-0.041}$	$0.19^{+0.08}_{-0.05}$	$0.27^{+0.03}_{-0.03}$
WASP-5	8343.440	0.589	0.216 ± 0.002	0.235 ± 0.012	0.64 ± 0.02	$-4.67^{+0.04}_{-0.04}{}^{+0.017}_{-0.018}$	$1.71^{+0.40}_{-0.32}$	$3.05^{+1.35}_{-1.35}$
WASP-6	8354.589	0.165	0.233 ± 0.002	0.264 ± 0.013	0.74 ± 0.02	$-4.66^{+0.03}_{-0.04}{}^{+0.019}_{-0.020}$	$1.59^{+0.34}_{-0.27}$	$11.0^{+7.00}_{-7.00}$
WASP-7	8419.337	0.983	0.184 ± 0.0004	0.182 ± 0.010	0.43 ± 0.02	$-4.82^{+0.05}_{-0.06}{}^{+0.004}_{-0.004}$	$3.49^{+0.87}_{-0.68}$	$2.40^{+0.80}_{-0.40}$
WASP-8	8352.402	0.199	0.249 ± 0.001	0.291 ± 0.013	0.70 ± 0.03	$-4.57^{+0.03}_{-0.03}{}^{+0.020}_{-0.020}$	$0.97^{+0.23}_{-0.18}$	$4.00^{+1.00}_{-1.00}$
WASP-15	8352.227	0.113	0.149 ± 0.001	0.124 ± 0.009	0.43 ± 0.02	$-5.31^{+0.13}_{-0.18}{}^{+0.021}_{-0.022}$	$11.02^{+1.27}_{-1.27}$	$2.40^{+0.60}_{-0.60}$
WASP-17	7998.290	0.659	0.155 ± 0.001	0.133 ± 0.010	0.4 ± 0.02	$-5.20^{+0.10}_{-0.13}{}^{+0.022}_{-0.023}$	$9.95^{+1.16}_{-1.61}$	$1.90^{+0.50}_{-0.50}$
WASP-18	8136.299	0.275	0.155 ± 0.001	0.133 ± 0.009	0.47 ± 0.02	$-5.18^{+0.10}_{-0.13}{}^{+0.013}_{-0.014}$	$9.73^{+1.29}_{-1.61}$	$0.63^{+0.95}_{-0.53}$
WASP-25	7952.288	0.131	0.206 ± 0.001	0.219 ± 0.011	0.65 ± 0.01	$-4.73^{+0.04}_{-0.04}{}^{+0.010}_{-0.010}$	$2.31^{+0.49}_{-0.40}$	$0.10^{+0.20}_{-0.00}$
WASP-26	8352.426	0.294	0.155 ± 0.002	0.133 ± 0.010	0.55 ± 0.02	$-5.20^{+0.10}_{-0.13}{}^{+0.033}_{-0.036}$	$9.97^{+1.14}_{-1.58}$	$4.00^{+1.40}_{-4.00}$
WASP-28	7952.558	0.957	0.159 ± 0.002	0.140 ± 0.010	0.52 ± 0.01	$-5.11^{+0.09}_{-0.11}{}^{+0.033}_{-0.036}$	$8.70^{+1.64}_{-1.62}$	$5.00^{+3.00}_{-2.00}$
WASP-32	7661.427	0.386	0.162 ± 0.004	0.144 ± 0.012	0.52 ± 0.03	$-5.07^{+0.09}_{-0.12}{}^{+0.057}_{-0.065}$	$7.97^{+1.94}_{-1.75}$	$2.12^{+0.94}_{-0.94}$
WASP-34	7819.320	0.600	0.156 ± 0.001	0.135 ± 0.009	0.66 ± 0.01	$-5.19^{+0.08}_{-0.10}{}^{+0.008}_{-0.008}$	$9.86^{+1.06}_{-1.30}$	$6.70^{+6.90}_{-4.50}$
WASP-36	7800.332	0.854	0.208 ± 0.004	0.221 ± 0.013	0.58 ± 0.02	$-4.69^{+0.04}_{-0.05}{}^{+0.023}_{-0.025}$	$1.87^{+0.51}_{-0.39}$	$1.80^{+2.70}_{-1.30}$
WASP-38	8237.541	0.249	0.156 ± 0.001	0.136 ± 0.009	0.51 ± 0.02	$-5.16^{+0.09}_{-0.12}{}^{+0.010}_{-0.010}$	$9.38^{+1.41}_{-1.58}$	$1.12^{+0.32}_{-0.32}$
WASP-43	6689.404	0.641	1.209 ± 0.008	1.889 ± 0.065	1.10 ± 0.06	$-4.17^{+0.09}_{-0.11}{}^{+0.085}_{-0.105}$	$0.05^{+0.07}_{-0.03}$	$0.50^{+0.07}_{-0.07}$
WASP-43	7558.218	0.671	1.190 ± 0.009	1.858 ± 0.064	1.10 ± 0.06	$-4.18^{+0.09}_{-0.11}{}^{+0.085}_{-0.105}$	$0.05^{+0.07}_{-0.03}$	$0.50^{+0.07}_{-0.07}$
WASP-44	7952.533	0.924	0.213 ± 0.003	0.231 ± 0.013	0.76 ± 0.05	$-4.76^{+0.05}_{-0.05}{}^{+0.037}_{-0.041}$	$2.65^{+0.73}_{-0.55}$	$0.90^{+1.00}_{-0.60}$
WASP-45	8298.568	0.015	0.197 ± 0.003	0.203 ± 0.012	0.91 ± 0.08	$-4.94^{+0.07}_{-0.08}{}^{+0.063}_{-0.074}$	$5.55^{+1.52}_{-1.20}$	$1.40^{+2.00}_{-1.00}$
WASP-46	7563.427	0.864	0.237 ± 0.012	0.270 ± 0.024	0.64 ± 0.05	$-4.58^{+0.06}_{-0.07}{}^{+0.055}_{-0.062}$	$1.02^{+0.51}_{-0.32}$	$1.40^{+0.40}_{-0.60}$
WASP-47	7569.668	0.061	0.161 ± 0.002	0.143 ± 0.010	0.73 ± 0.02	$-5.13^{+0.07}_{-0.08}{}^{+0.026}_{-0.028}$	$8.88^{+1.18}_{-1.18}$	$6.50^{+2.60}_{-1.20}$
WASP-50	8353.536	0.560	0.442 ± 0.004	0.612 ± 0.023	0.73 ± 0.01	$-4.18^{+0.02}_{-0.02}{}^{+0.015}_{-0.015}$	$0.05^{+0.01}_{-0.01}$	$8.10^{+1.50}_{-1.30}$
WASP-51(5)	6698.410	0.883	0.168 ± 0.001	0.155 ± 0.010	0.49 ± 0.01	$-4.98^{+0.07}_{-0.08}{}^{+0.006}_{-0.006}$	$6.23^{+1.46}_{-1.20}$	$1.00^{+0.80}_{-0.50}$
WASP-52	8353.461	0.914	0.442 ± 0.004	0.612 ± 0.023	0.90 ± 0.04	$-4.38^{+0.05}_{-0.06}{}^{+0.052}_{-0.059}$	$0.26^{+0.15}_{-0.09}$	$0.40^{+0.30}_{-0.20}$
WASP-55	7979.233	0.899	0.162 ± 0.002	0.145 ± 0.010	0.55 ± 0.01	$-5.08^{+0.08}_{-0.10}{}^{+0.024}_{-0.026}$	$8.02^{+1.62}_{-1.48}$	$3.00^{+5.00}_{-2.00}$
WASP-62	7802.402	0.303	0.194 ± 0.001	0.198 ± 0.011	0.51 ± 0.02	$-4.75^{+0.04}_{-0.05}{}^{+0.007}_{-0.007}$	$2.55^{+0.63}_{-0.49}$	$0.70^{+0.40}_{-0.30}$
WASP-63	7766.531	0.389	0.142 ± 0.002	0.111 ± 0.010	0.68 ± 0.02	$-5.50^{+0.16}_{-0.25}{}^{+0.067}_{-0.079}$	> 11.0	$6.00^{+5.00}_{-3.00}$
WASP-63	7979.664	0.071	0.136 ± 0.001	0.102 ± 0.009	0.68 ± 0.02	$-5.73^{+0.23}_{-0.50}{}^{+0.059}_{-0.068}$	> 11.0	$6.00^{+5.00}_{-3.00}$
WASP-65	7792.402	0.566	0.185 ± 0.003	0.183 ± 0.011	0.67 ± 0.03	$-4.88^{+0.05}_{-0.06}{}^{+0.026}_{-0.028}$	$4.47^{+1.02}_{-0.82}$	$1.72^{+1.26}_{-0.76}$
WASP-67	7845.615	0.028	0.197 ± 0.002	0.203 ± 0.011	0.77 ± 0.03	$-4.85^{+0.04}_{-0.04}{}^{+0.020}_{-0.021}$	$3.95^{+0.72}_{-0.60}$	$2.00^{+1.60}_{-1.00}$
WASP-67	7965.286	0.962	0.213 ± 0.002	0.231 ± 0.012	0.77 ± 0.03	$-4.76^{+0.04}_{-0.04}{}^{+0.021}_{-0.023}$	$2.73^{+0.55}_{-0.45}$	$2.00^{+1.60}_{-1.00}$
WASP-69	7996.327	0.027	0.415 ± 0.001	0.568 ± 0.021	1.04 ± 0.02	$-4.61^{+0.04}_{-0.04}{}^{+0.036}_{-0.039}$	$1.25^{+0.34}_{-0.26}$	$1.10^{+0.15}_{-0.15}$
WASP-70	8007.434	0.613	0.145 ± 0.001	0.117 ± 0.009	0.65 ± 0.03	$-5.42^{+0.13}_{-0.19}{}^{+0.036}_{-0.040}$	> 11.0	$7.00^{+3.00}_{-3.00}$
WASP-72	6606.307	0.332	0.149 ± 0.001	0.124 ± 0.009	0.48 ± 0.03	$-5.30^{+0.13}_{-0.18}{}^{+0.027}_{-0.029}$	$10.99^{+-0.29}_{-1.30}$	$3.70^{+4.00}_{-1.90}$
WASP-73	7717.291	0.665	0.144 ± 0.001	0.115 ± 0.009	0.56 ± 0.04	$-5.47^{+0.16}_{-0.26}{}^{+0.030}_{-0.033}$	> 11.0	$4.55^{+1.85}_{-1.85}$
WASP-74	7934.457	0.788	0.155 ± 0.001	0.134 ± 0.009	0.61 ± 0.04	$-5.20^{+0.09}_{-0.12}{}^{+0.014}_{-0.014}$	$9.96^{+1.09}_{-1.42}$	$4.20^{+1.60}_{-2.00}$
WASP-75	8348.412	0.791	0.151 ± 0.001	0.126 ± 0.009	0.54 ± 0.03	$-5.28^{+0.12}_{-0.16}{}^{+0.025}_{-0.027}$	$10.84^{+0.17}_{-1.35}$	$1.69^{+1.58}_{-0.87}$
WASP-76	7695.414	0.163	0.144 ± 0.003	0.115 ± 0.010	0.52 ± 0.03	$-5.47^{+0.18}_{-0.32}{}^{+0.089}_{-0.113}$	> 11.0	$5.30^{+6.10}_{-2.90}$
WASP-76	8008.492	0.145	0.145 ± 0.001	0.116 ± 0.009	0.52 ± 0.03	$-5.44^{+0.16}_{-0.25}{}^{+0.023}_{-0.025}$	> 11.0	$5.30^{+6.10}_{-2.90}$

Target[a]	MJD −2450 K	Phase [ϕ]	$S_{\rm RSS}$	$S_{\rm MW}$	$B-V$[b]	$\log(R'_{\rm HK})$	Act Age [c] [Gyr]	Gyro age[d] [Gyr]
WASP-77	7994.540	0.796	0.223 ± 0.001	0.247 ± 0.012	0.69 ± 0.01	$-4.67^{+0.03}_{-0.03}{}^{+0.011}_{-0.011}$	$1.69^{+0.34}_{-0.28}$	$1.00^{+0.50}_{-0.30}$
WASP-78	8354.574	0.559	0.143 ± 0.001	0.113 ± 0.009	0.49 ± 0.04	$-5.51^{+0.18}_{-0.33}{}^{+0.046}_{-0.052}$	> 11.0	$1.37^{+1.91}_{-0.78}$
WASP-79	7747.487	0.315	0.184 ± 0.001	0.182 ± 0.010	0.42 ± 0.02	$-4.82^{+0.05}_{-0.06}{}^{+0.007}_{-0.007}$	$3.51^{+0.88}_{-0.69}$	$0.60^{+0.00}_{-0.00}$
WASP-80	8239.635	0.152	1.285 ± 0.010	2.017 ± 0.070	1.26 ± 0.05	$-4.41^{+0.08}_{-0.10}{}^{+0.080}_{-0.099}$	$0.31^{+0.32}_{-0.15}$	$0.10^{+0.03}_{-0.02}$
WASP-80	8348.424	0.613	1.178 ± 0.011	1.838 ± 0.065	1.26 ± 0.05	$-4.45^{+0.08}_{-0.10}{}^{+0.080}_{-0.099}$	$0.42^{+0.40}_{-0.19}$	$0.10^{+0.03}_{-0.02}$
WASP-82	8157.333	0.914	0.143 ± 0.001	0.114 ± 0.009	0.44 ± 0.02	$-5.50^{+0.18}_{-0.31}{}^{+0.024}_{-0.026}$	> 11.0	
WASP-87 A	7586.271	0.617	0.172 ± 0.001	0.161 ± 0.010	0.41 ± 0.03	$-4.95^{+0.06}_{-0.07}{}^{+0.014}_{-0.015}$	$5.59^{+1.38}_{-1.12}$	1.68
WASP-88	7563.439	0.763	0.163 ± 0.003	0.146 ± 0.011	0.44 ± 0.03	$-5.05^{+0.09}_{-0.11}{}^{+0.043}_{-0.047}$	$7.59^{+1.85}_{-1.61}$	$3.55^{+1.80}_{-1.80}$
WASP-89	7884.560	0.800	0.429 ± 0.005	0.590 ± 0.023	0.94 ± 0.04	$-4.44^{+0.06}_{-0.07}{}^{+0.056}_{-0.065}$	$0.41^{+0.24}_{-0.14}$	$1.30^{+1.50}_{-0.80}$
WASP-90	7569.526	0.623	0.149 ± 0.001	0.123 ± 0.009	0.46 ± 0.03	$-5.32^{+0.13}_{-0.19}{}^{+0.031}_{-0.033}$	$11.07^{+1.22}_{-1.22}$	$4.40^{+8.40}_{-24.00}$
WASP-91	8330.509	0.365	0.178 ± 0.002	0.172 ± 0.011	0.96 ± 0.04	$-5.08^{+0.04}_{-0.05}{}^{+0.033}_{-0.036}$	$8.01^{+0.87}_{-0.83}$	$0.86^{+0.30}_{-0.30}$
WASP-94 A	7565.428	0.879	0.164 ± 0.001	0.147 ± 0.010	0.54 ± 0.03	$-5.05^{+0.07}_{-0.09}{}^{+0.011}_{-0.011}$	$7.55^{+1.56}_{-1.38}$	$4.90^{+1.72}_{-1.72}$
WASP-94 B	7565.428		0.162 ± 0.001	0.145 ± 0.010	0.57 ± 0.03	$-5.08^{+0.08}_{-0.09}{}^{+0.013}_{-0.014}$	$8.16^{+1.53}_{-1.42}$	$4.90^{+1.72}_{-1.72}$
WASP-95	7965.376	0.696	0.181 ± 0.001	0.176 ± 0.010	0.63 ± 0.05	$-4.90^{+0.05}_{-0.06}{}^{+0.020}_{-0.021}$	$4.70^{+1.07}_{-0.87}$	$2.40^{+1.70}_{-1.00}$
WASP-95	7937.461	0.918	0.185 ± 0.001	0.184 ± 0.010	0.63 ± 0.05	$-4.86^{+0.05}_{-0.06}{}^{+0.021}_{-0.022}$	$4.09^{+0.96}_{-0.77}$	$8.00^{+26.00}_{-8.00}$
WASP-96	7952.493	0.687	0.148 ± 0.002	0.121 ± 0.010	0.73 ± 0.06	$-5.32^{+0.10}_{-0.13}{}^{+0.049}_{-0.055}$	$11.07^{+0.85}_{-0.85}$	$11.90^{+16.00}_{-8.30}$
WASP-97	7650.377	0.819	0.150 ± 0.001	0.124 ± 0.009	0.69 ± 0.04	$-5.30^{+0.10}_{-0.13}{}^{+0.020}_{-0.021}$	$10.97^{+0.10}_{-0.96}$	$1.40^{+1.10}_{-0.60}$
WASP-99	8132.370	0.575	0.155 ± 0.0004	0.133 ± 0.009	0.54 ± 0.03	$-5.19^{+0.10}_{-0.13}{}^{+0.010}_{-0.011}$	$9.91^{+1.16}_{-1.54}$	
WASP-100	8089.350	0.687	0.186 ± 0.001	0.185 ± 0.010	0.35 ± 0.02	$-4.84^{+0.05}_{-0.06}{}^{+0.018}_{-0.019}$	$3.84^{+0.98}_{-0.76}$	$0.90^{+1.30}_{-0.40}$
WASP-101	8008.631	0.246	0.200 ± 0.001	0.209 ± 0.011	0.49 ± 0.03	$-4.70^{+0.04}_{-0.04}{}^{+0.006}_{-0.006}$	$2.00^{+0.49}_{-0.38}$	$4.00^{+1.00}_{-1.00}$
WASP-103	6779.562	0.701	0.224 ± 0.003	0.248 ± 0.013	0.54 ± 0.04	$-4.59^{+0.04}_{-0.04}{}^{+0.021}_{-0.022}$	$1.05^{+0.29}_{-0.22}$	$4.00^{+1.00}_{-1.00}$
WASP-103	6779.566	0.706	0.221 ± 0.003	0.244 ± 0.013	0.54 ± 0.04	$-4.60^{+0.04}_{-0.04}{}^{+0.021}_{-0.022}$	$1.12^{+0.31}_{-0.23}$	$4.00^{+1.00}_{-1.00}$
WASP-103	6779.573	0.713	0.221 ± 0.003	0.243 ± 0.013	0.54 ± 0.04	$-4.60^{+0.04}_{-0.04}{}^{+0.021}_{-0.022}$	$1.14^{+0.31}_{-0.24}$	$4.00^{+1.00}_{-1.00}$
WASP-103	6780.489	0.703	0.227 ± 0.003	0.253 ± 0.013	0.54 ± 0.04	$-4.57^{+0.04}_{-0.04}{}^{+0.021}_{-0.022}$	$0.96^{+0.27}_{-0.20}$	$4.00^{+1.00}_{-1.00}$
WASP-103	6780.493	0.707	0.227 ± 0.003	0.254 ± 0.013	0.54 ± 0.04	$-4.57^{+0.04}_{-0.04}{}^{+0.020}_{-0.021}$	$0.95^{+0.26}_{-0.20}$	$4.00^{+1.00}_{-1.00}$
WASP-103	6780.497	0.712	0.233 ± 0.003	0.263 ± 0.013	0.54 ± 0.04	$-4.54^{+0.04}_{-0.04}{}^{+0.020}_{-0.021}$	$0.81^{+0.22}_{-0.17}$	$4.00^{+1.00}_{-1.00}$
WASP-103	6780.501	0.716	0.230 ± 0.003	0.258 ± 0.013	0.54 ± 0.04	$-4.56^{+0.04}_{-0.04}{}^{+0.020}_{-0.021}$	$0.88^{+0.24}_{-0.18}$	$4.00^{+1.00}_{-1.00}$
WASP-103	6780.535	0.753	0.220 ± 0.009	0.241 ± 0.019	0.54 ± 0.04	$-4.61^{+0.06}_{-0.06}{}^{+0.045}_{-0.050}$	$1.18^{+0.50}_{-0.34}$	$4.00^{+1.00}_{-1.00}$
WASP-103	6790.473	0.490	0.229 ± 0.003	0.256 ± 0.013	0.54 ± 0.04	$-4.56^{+0.04}_{-0.04}{}^{+0.020}_{-0.021}$	$0.92^{+0.25}_{-0.19}$	$4.00^{+1.00}_{-1.00}$
WASP-103	6790.478	0.496	0.229 ± 0.003	0.257 ± 0.013	0.54 ± 0.04	$-4.56^{+0.04}_{-0.04}{}^{+0.020}_{-0.021}$	$0.91^{+0.25}_{-0.19}$	$4.00^{+1.00}_{-1.00}$
WASP-103	6790.482	0.500	0.235 ± 0.003	0.267 ± 0.013	0.54 ± 0.04	$-4.53^{+0.04}_{-0.04}{}^{+0.020}_{-0.021}$	$0.76^{+0.21}_{-0.16}$	$4.00^{+1.00}_{-1.00}$
WASP-103	6790.487	0.505	0.237 ± 0.003	0.270 ± 0.013	0.54 ± 0.04	$-4.53^{+0.03}_{-0.04}{}^{+0.019}_{-0.020}$	$0.73^{+0.20}_{-0.15}$	$4.00^{+1.00}_{-1.00}$
WASP-103	6790.502	0.521	0.227 ± 0.003	0.254 ± 0.013	0.54 ± 0.04	$-4.57^{+0.04}_{-0.04}{}^{+0.020}_{-0.021}$	$0.95^{+0.26}_{-0.20}$	$4.00^{+1.00}_{-1.00}$
WASP-103	6790.507	0.527	0.234 ± 0.003	0.264 ± 0.013	0.54 ± 0.04	$-4.54^{+0.04}_{-0.04}{}^{+0.019}_{-0.020}$	$0.80^{+0.22}_{-0.16}$	$4.00^{+1.00}_{-1.00}$
WASP-103	6790.511	0.531	0.225 ± 0.003	0.251 ± 0.013	0.54 ± 0.04	$-4.58^{+0.04}_{-0.04}{}^{+0.019}_{-0.020}$	$1.00^{+0.27}_{-0.20}$	$4.00^{+1.00}_{-1.00}$
WASP-103	6790.516	0.537	0.231 ± 0.003	0.260 ± 0.013	0.54 ± 0.04	$-4.55^{+0.04}_{-0.04}{}^{+0.019}_{-0.020}$	$0.86^{+0.23}_{-0.18}$	$4.00^{+1.00}_{-1.00}$
WASP-104	7765.547	0.428	0.149 ± 0.002	0.123 ± 0.010	0.77 ± 0.05	$-5.28^{+0.08}_{-0.10}{}^{+0.031}_{-0.033}$	$10.86^{+0.32}_{-0.86}$	$3.00^{+2.00}_{-2.00}$
WASP-104	7910.228	0.848	0.143 ± 0.002	0.113 ± 0.010	0.77 ± 0.05	$-5.38^{+0.11}_{-0.14}{}^{+0.054}_{-0.061}$	> 11.0	$3.00^{+2.00}_{-2.00}$
WASP-105	8353.446	0.710	0.166 ± 0.002	0.152 ± 0.010	0.91 ± 0.06	$-5.11^{+0.05}_{-0.06}{}^{+0.032}_{-0.035}$	$8.68^{+0.89}_{-0.87}$	$2.04^{+4.40}_{-4.40}$
WASP-108	8237.522	0.630	0.161 ± 0.002	0.143 ± 0.010	0.57 ± 0.04	$-5.10^{+0.08}_{-0.10}{}^{+0.031}_{-0.033}$	$8.41^{+1.62}_{-1.54}$	$4.60^{+1.90}_{-1.90}$
WASP-109	7918.456	0.193	0.209 ± 0.001	0.224 ± 0.011	0.41 ± 0.03	$-4.65^{+0.04}_{-0.04}{}^{+0.010}_{-0.010}$	$1.53^{+0.38}_{-0.30}$	3.32
WASP-111	8417.365	0.718	0.167 ± 0.001	0.154 ± 0.010	0.46 ± 0.04	$-4.99^{+0.07}_{-0.08}{}^{+0.007}_{-0.007}$	$6.35^{+1.49}_{-1.23}$	$2.60^{+0.60}_{-0.60}$
WASP-114	8352.439	0.765	0.161 ± 0.003	0.142 ± 0.011	0.59 ± 0.04	$-5.11^{+0.09}_{-0.11}{}^{+0.046}_{-0.052}$	$8.60^{+1.66}_{-1.62}$	$0.98^{+0.51}_{-0.31}$

Target[a]	MJD −2450 K	Phase [ϕ]	$S_{\rm RSS}$	$S_{\rm MW}$	$B-V$[b]	$\log(R'_{\rm HK})$	Act Age [c] [Gyr]	Gyro age[d] [Gyr]
WASP-117	8354.487	0.692	0.163 ± 0.001	0.146 ± 0.010	0.54 ± 0.03	$-5.06^{+0.08}_{-0.09}{}^{+0.010}_{-0.011}$	$7.76^{+1.58}_{-1.42}$	$3.83^{+1.54}_{-1.54}$
WASP-118	7980.589	0.800	0.170 ± 0.001	0.158 ± 0.010	0.47 ± 0.03	$-4.96^{+0.06}_{-0.07}{}^{+0.008}_{-0.008}$	$5.80^{+1.40}_{-1.13}$	$1.17^{+5.72}_{-0.75}$
WASP-119	8093.293	0.351	0.179 ± 0.002	0.173 ± 0.011	0.68 ± 0.03	$-4.93^{+0.05}_{-0.06}{}^{+0.022}_{-0.023}$	$5.38^{+1.11}_{-0.92}$	$8.00^{+2.50}_{-2.50}$
WASP-120	8354.552	0.167	0.162 ± 0.001	0.145 ± 0.010	0.44 ± 0.03	$-5.06^{+0.08}_{-0.09}{}^{+0.010}_{-0.010}$	$7.67^{+1.63}_{-1.45}$	$2.60^{+0.50}_{-0.50}$
WASP-122	8419.516	0.868	0.157 ± 0.001	0.137 ± 0.010	0.68 ± 0.04	$-5.18^{+0.08}_{-0.10}{}^{+0.013}_{-0.014}$	$9.67^{+1.10}_{-1.27}$	$5.10^{+0.80}_{-0.80}$
WASP-123	7952.548	0.897	0.150 ± 0.001	0.125 ± 0.009	0.66 ± 0.04	$-5.30^{+0.10}_{-0.14}{}^{+0.021}_{-0.022}$	$10.96^{+0.06}_{-1.04}$	$6.90^{+1.40}_{-1.40}$
WASP-126	7801.292	0.992	0.149 ± 0.001	0.123 ± 0.009	0.64 ± 0.03	$-5.33^{+0.11}_{-0.15}{}^{+0.021}_{-0.022}$	$11.12^{+90}_{-0.90}$	$6.40^{+1.60}_{-1.60}$
WASP-127	7782.485	0.749	0.155 ± 0.001	0.133 ± 0.009	0.61 ± 0.03	$-5.20^{+0.09}_{-0.12}{}^{+0.013}_{-0.014}$	$10.06^{+1.02}_{-1.40}$	$11.41^{+0.80}_{-0.80}$
WASP-130	8239.519	0.135	0.170 ± 0.001	0.158 ± 0.010	0.71 ± 0.03	$-5.03^{+0.06}_{-0.06}{}^{+0.012}_{-0.012}$	$7.11^{+1.17}_{-1.04}$	$4.05^{+3.85}_{-3.85}$
WASP-131	7819.445	0.037	0.153 ± 0.001	0.130 ± 0.009	0.55 ± 0.03	$-5.23^{+0.10}_{-0.13}{}^{+0.012}_{-0.012}$	$10.38^{+0.81}_{-1.46}$	$4.05^{+3.85}_{-3.85}$
WASP-131	7907.436	0.571	0.153 ± 0.001	0.130 ± 0.009	0.55 ± 0.03	$-5.23^{+0.10}_{-0.13}{}^{+0.012}_{-0.012}$	$10.38^{+0.81}_{-1.46}$	$7.50^{+2.50}_{-2.50}$
WASP-132	7979.276	0.584	0.387 ± 0.002	0.520 ± 0.020	1.03 ± 0.05	$-4.63^{+0.07}_{-0.08}{}^{+0.064}_{-0.076}$	$1.40^{+0.73}_{-0.45}$	$2.20^{+0.30}_{-0.30}$
WASP-138	8354.563	0.834	0.154 ± 0.001	0.132 ± 0.010	0.47 ± 0.03	$-5.19^{+0.10}_{-0.13}{}^{+0.021}_{-0.022}$	$9.83^{+1.24}_{-1.62}$	$3.44^{+0.93}_{-0.93}$
WASP-139	7948.631	0.908	0.159 ± 0.002	0.140 ± 0.010	0.81 ± 0.04	$-5.15^{+0.06}_{-0.06}{}^{+0.017}_{-0.017}$	$9.27^{+0.91}_{-0.94}$	$0.50^{+0.40}_{-0.30}$
WASP-140	8157.343	0.798	0.356 ± 0.002	0.469 ± 0.018	0.80 ± 0.04	$-4.39^{+0.04}_{-0.05}{}^{+0.038}_{-0.042}$	$0.29^{+0.12}_{-0.08}$	$1.60^{+1.40}_{-0.90}$
WASP-142	7891.308	0.388	0.141 ± 0.003	0.110 ± 0.010	0.57 ± 0.04	$-5.58^{+0.21}_{-0.43}{}^{+0.104}_{-0.137}$	> 11.0	2.00
WASP-156	8413.516	0.838	0.179 ± 0.002	0.174 ± 0.011	0.97 ± 0.03	$-5.08^{+0.04}_{-0.05}{}^{+0.028}_{-0.030}$	$8.09^{+0.79}_{-0.75}$	$0.58^{+0.51}_{-0.31}$

[a] Commonly used alternative names: (1) HIP 116454 (2) HD 3167 (3) GJ 9837 (4) EPIC 212756297 (5) HAT-P-30.

[b] Calculated with Equation 3 of Sekiguchi & Fukugita (2000).

[c] Calculated using Equation 3 of Mamajek & Hillenbrand (2008).

[d] Calculated with Equation 3 from Barnes (2007).

3.6 Comparison to published activity values

Targets with published activity values are listed in Table 3.2, and compared to OU-SALT $\log(R'_{\rm HK})$ measurements in Figure 3.7. Approximately 89 % of targets are within $|\Delta \log(R'_{\rm HK})| < 0.25$, with only WASP-15 (Triaud et al., 2010) and WASP-50 (Gillon et al., 2011) showing larger disparities. WASP-50 is a highly active target, so significant activity variation is expected. For WASP-15, $\Delta \log(R'_{\rm HK}) = -0.45$, meaning the target is classified as sub-basal by OU-SALT but super-basal by its alternative activity value. This may be due to genuine magnetic variability or measurement error.

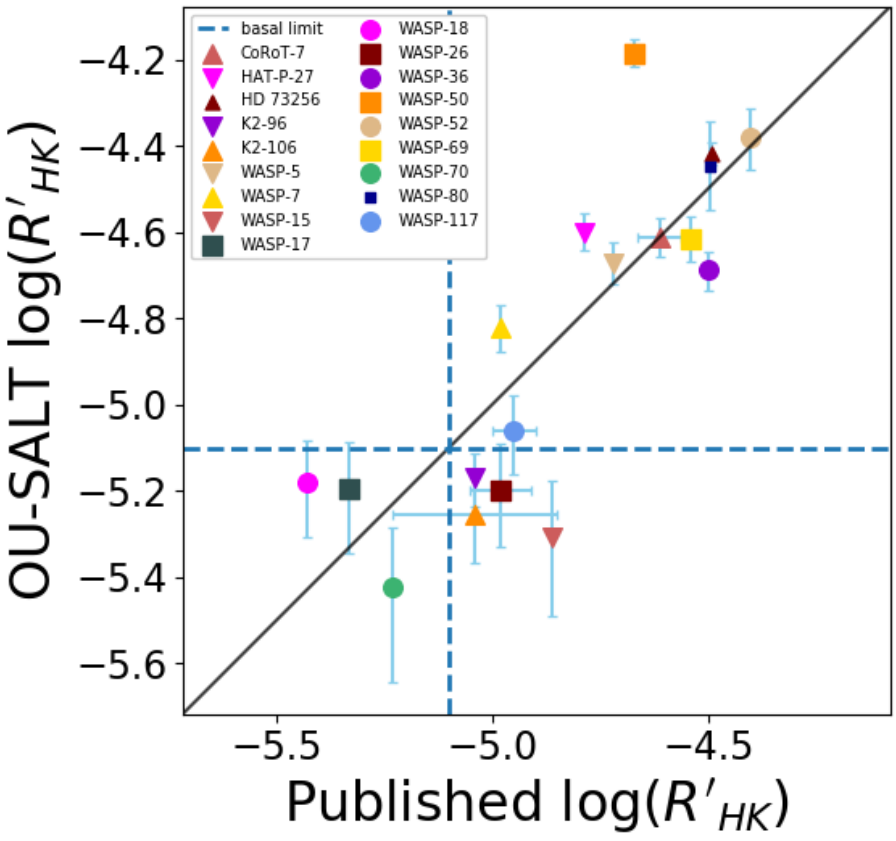

FIGURE 3.7: Comparisons of OU-SALT activity metrics and published values. Each target is represented by a unique symbol. A line of equivalence is plotted in black. The sub-basal limit ($\log(R'_{\rm HK})$ = -5.1) is shown by dashed blue lines.

TABLE 3.2: OU-SALT activity values compared to published $\log(R'_{HK})$ metrics.

Target	OU-SALT $\log(R'_{HK})$	Published $\log(R'_{HK})$[a]
CoRoT-7	$-4.61^{+0.03}_{-0.03}$	-4.61±0.05 (1)
HAT-P-27	$-4.60^{+0.03}_{-0.03}$	-4.79 (2)
HD 73256	$-4.42^{+0.03}_{-0.03}$	-4.49 (3)
K2-96	$-5.17^{+0.06}_{-0.07}$	-5.04 (4)
K2-106	$-5.25^{+0.09}_{-0.11}$	-5.04±0.19 (5)
WASP-5	$-4.67^{+0.04}_{-0.04}$	-4.72 (6)
WASP-7	$-4.82^{+0.05}_{-0.06}$	-4.98 (7)
WASP-15	$-5.31^{+0.13}_{-0.18}$	-4.86 (6)
WASP-17	$-5.20^{+0.10}_{-0.13}$	-5.33 (8)
WASP-18	$-5.18^{+0.10}_{-0.13}$	-5.43 (8)
WASP-26	$-5.20^{+0.10}_{-0.13}$	-4.98±0.07 (9)
WASP-36	$-4.69^{+0.04}_{-0.05}$	-4.50 (10)
WASP-50	$-4.18^{+0.02}_{-0.02}$	-4.67 (11)
WASP-52	$-4.38^{+0.05}_{-0.06}$	-4.40 (12)
WASP-69	$-4.61^{+0.04}_{-0.04}$	-4.54 (9)
WASP-70	$-5.42^{+0.13}_{-0.19}$	-5.23 (9)
WASP-80	$-4.45^{+0.08}_{-0.10}$	-4.50 (13)
WASP-117	$-5.06^{+0.08}_{-0.09}$	-4.95±0.05 (14)

[a] References: (1) Queloz et al. (2009); (2) Béky et al. (2011); (3) Henry et al. (1996); (4) Christiansen et al. (2017); (5) Guenther et al. (2017); (6) Triaud et al. (2010); (7) Albrecht et al. (2012); (8) Knutson et al. (2010); (9) Anderson et al. (2011); (10) Smith et al. (2012); (11) Gillon et al. (2011); (12) Hébrard et al. (2013); (13) Mancini et al. (2014); (14) Lendl et al. (2014).

3.6.1 Systematic offset

Figure 3.8 shows a histogram of residuals of the published activity values subtracted from OU-SALT $\log(R'_{\rm HK})$. The mean of the fitted Gaussian curve is $\mu = 9.4 \times 10^{-3}$ (with $\sigma = 0.2$), indicating that there is no significant systematic offset. I statistically tested for systematic offset using the paired Student's t-test (Student, 1908). The test statistic $|t_e|$ value for the targets detailed in Table 3.2 is 0.19. This is less than the $N - 1$, 50% confidence level critical value ($t_e = 0.69$)[5] for this sample size. After removing WASP-15 and WASP-50 from the sample, $|t_e| = 0.35$. The null hypothesis, that there is no significant systematic offset between these data sets, is accepted.

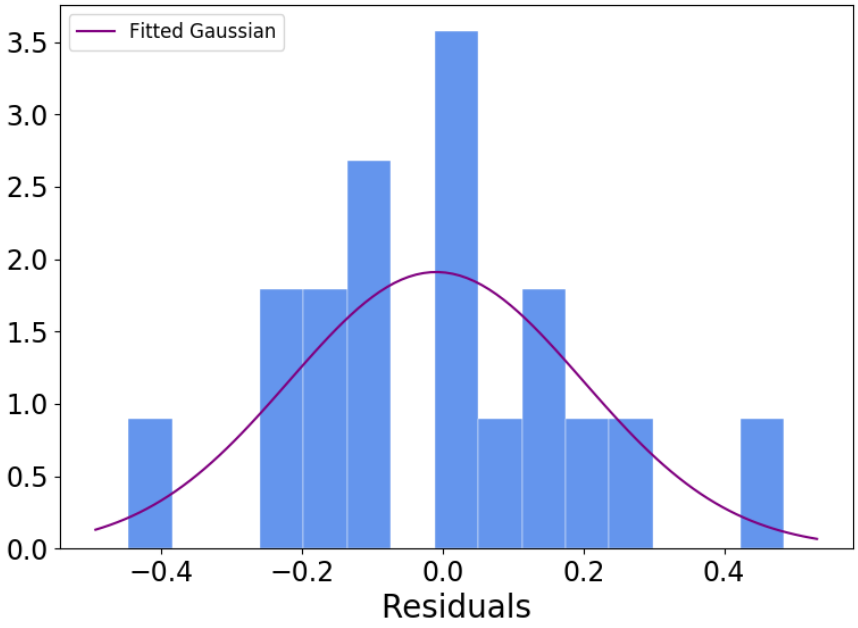

FIGURE 3.8: A histogram of residuals between OU-SALT and published $\log(R'_{\rm HK})$ activity values. No systematic offset is apparent.

[5] http://www.sjsu.edu/faculty/gerstman/StatPrimer/t-table.pdf

3.7 Summary

The Robert Stobie Spectrograph (RSS) on the Southern African Large Telescope (SALT) has been calibrated to produce precise $\log(R'_{\rm HK})$ measurements of stars hosting transiting planets. Our RSS set-up has been detailed, and the data reduction process outlined. This included image trimming, wavelength calibration, flat-fielding, cosmic ray removal, as well as background and spectral extraction. A description of some of the specific challenges faced was provided. The conversion to the S-index and ultimately $\log(R'_{\rm HK})$ has been detailed. Generating this widely-adopted metric allows comparison of our work to other studies.

The OU-SALT homogeneous activity database of 104 close-orbiting planet hosts was presented with both internal and calibration uncertainties. The median internal uncertainty is ~ 3.3 times smaller than the external error. One of the key advantages of this study is that internal errors may be used in analysis, providing greater scientific resolution. Our metrics were compared to published activity values of 18 hosts. A paired Student's t-test statistic value of 0.19 indicates that the null hypothesis (of no significant offset between data sets) may be accepted.

The following chapter will address the demographics of the OU-SALT sample and what scientific information may drawn from our unique dataset.

Collaboration details

The OU-SALT survey was started by Carole Haswell. Calibration of the Robert Stobie Spectrograph was undertaken by Staab et al. (2017). Daniel Staab assisted in bringing me up-to-speed with this project. I have planned and submitted observing proposals for SALT semesters since 2016. Observations have been carried out by resident SALT astronomers using the set-up detailed. I have generated activity measurements for a 104 planet hosts, improved on previous analyses, and introduced efficiencies into the reduction process. This work has been undertaken in regular consultation with Carole Haswell and John Barnes.

Chapter 4

OU-SALT systems

This chapter explores the systems in the OU-SALT sample. It begins with a description of the target selection process in Section 4.1. The stellar evolutionary status of hosts is defined using the Hertzsprung-Russell (HR) diagram, Gaia DR2 data and isochrones in Section 4.2. A discussion of the host star age ensues in Section 4.3. The demographics of the planet population are considered in Section 4.4, along with aspects of the physical environment experienced by planets—including incident flux and Roche potential. The strength of interactions between stars and planets are considered in Section5.5. Time-series data is explored for systems with multiple observations in Section 4.6. I undertake a novel search for orbitally-modulated SPI in Section 4.7. In Section 4.8, I provide a summary of findings and details of collaboration.

4.1 Target selection

The goal of the OU-SALT survey is to generate a homogeneous activity database for a large sample of confirmed exoplanet hosts. I have focused my observations on close-orbiting hosts with a view to studying atmospheric escape and star-planet interactions (SPI) in the most extreme planetary systems. Targets were selected from the *Exoplanet.eu*[1] catalogue using the following criteria:

- declination $< +12$ deg so as to be observable by SALT;
- visual magnitude of $V < 13$ so that targets are bright enough to be observed in poor conditions with 400 s exposure times and signal-to-noise ratio (SNR) > 15 in Ca II H & K cores;
- effective temperature of 4145 K $< T_{\rm eff} <$ 6800 K (corresponding to main sequence spectral types F3-K7), as $\log(R'_{\rm HK})$ is well-calibrated within this range (Noyes et al., 1984; Isaacson & Fischer, 2010);
- transiting systems were chosen as science targets so that ablated planetary material lies in the plane of observation; and
- semi-major axis $a < 0.1$ AU to observe systems with the closest-in planets.

These criteria have allowed us to consistently produce high quality activity measurements of extreme planetary systems using the instrumentation at our disposal. Several exceptions were made:

- two non-transiting targets were observed: HD 73256 (Udry et al., 2003) by design and WASP-94 B (Neveu-VanMalle et al., 2014) opportunistically;
- several fainter targets of interest with $V > 13$ were observed with longer exposures; and
- the 36 M_J brown dwarf EPIC 219388192 was observed.

[1] http://exoplanet.eu/catalog/

4.2 Stellar evolutionary status

The activity distributions of main sequence (MS), sub-giant (SG) and giant stars are distinct (Mittag et al., 2013; Staab et al., 2017). It is therefore necessary to distinguish between these stellar classes in the OU-SALT sample.

4.2.1 Distance to target

The distance between the Earth and each target is required to calculate absolute magnitude (M_V). Gaia Data Release 2 (DR2) provides parallaxes of unprecedented accuracy (Brown et al., 2018). However, caution is required when working with DR2 data as individual parallaxes for stars beyond 1 kpc are unreliable (Luri et al., 2018). The use of Bayesian methods to estimate distances from DR2 parallaxes, especially for more distant objects, is encouraged (Arenou et al., 2018; Lindegren et al., 2018; Luri et al., 2018). Bailer-Jones et al. (2018) (BJ18) have undertaken this task for the majority of Gaia DR2 stars with parallaxes.

All OU-SALT targets are all located within 1 kpc, so inverting Gaia DR2 parallaxes should be safe. Due to the more rigorous Bayesian approach to distance calculation, BJ18 distances were nevertheless used in this work. Accessed via the Gaia archive[2], BJ18 distances place OU-SALT targets $\sim$3 pc closer on average than inverted parallaxes. Both approaches have percentage errors of approximately 2.1 %. Figure 4.1 shows visual magnitude plotted against BJ18 distances for OU-SALT targets. The uncertainties on measurements become more marked beyond 400 pc. WASP-103 (Gillon et al., 2014) has exceptionally large uncertainties in Figure 4.1, with $d = 833^{+178.43}_{-127.76}$ pc. This distance is considerably further than the Hipparcos distance (470 ± 35 pc), with significantly larger error bars. Southworth & Evans (2016) note the presence of a faint star near WASP-103. Flux from this companion, or perhaps its influence on the motion of WASP-103, may have effected Gaia's measurements, leading to larger uncertainties.

[2] https://gea.esac.esa.int/archive/

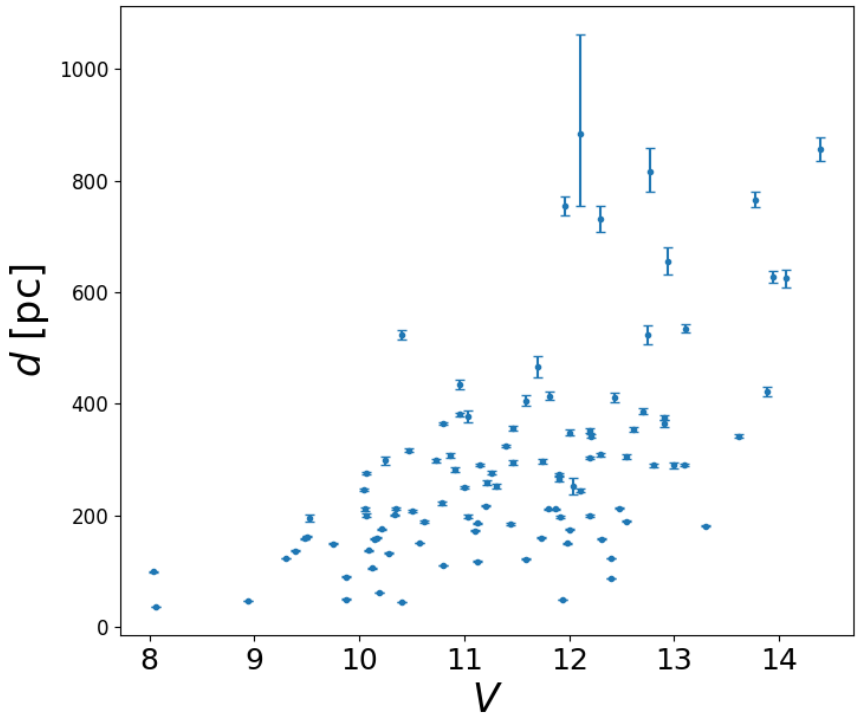

FIGURE 4.1: Gaia DR2 distances from BJ18 plotted against V magnitude. Uncertainties increase with distance. The target with exceptionally large error bars is WASP-103.

4.2.2 Hertzsprung-Russell diagrams

Colour-magnitude diagrams of the OU-SALT sample are plotted in Figure 4.2, with metallicity and activity colour categories indicated. While the majority of targets are clustered around the empirical main sequence of Wright (2005a) (shown as a solid black line), a number of hosts have evolved off the main sequence towards the subgiant branch, shown as a solid red line (Allen, 1973). Absolute magnitude is effected by several stellar parameters. Notably, metallicity alters the opacity of stellar material. Wright (2004) predicted that a change in metallicity of 0.3 dex would cause a corresponding shift in absolute magnitude of 0.45 mag (shown as dashed black lines in Figure 4.2). Wright (2004) defined ΔM_V, being the vertical difference between absolute magnitude and the empirically defined MS. This term is plotted against OU-SALT $\log(R'_{\rm HK})$ values in Figure 4.3, with metallicity categories indicated. The plot illustrates that, if metallicity is not taken into account when classifying targets, incorrect classifications ensue. For example, higher metallicity targets ([Fe/H]>0.3) with $0.45 < \Delta M_V < 0.90$ would be incorrectly classified as evolved stars.

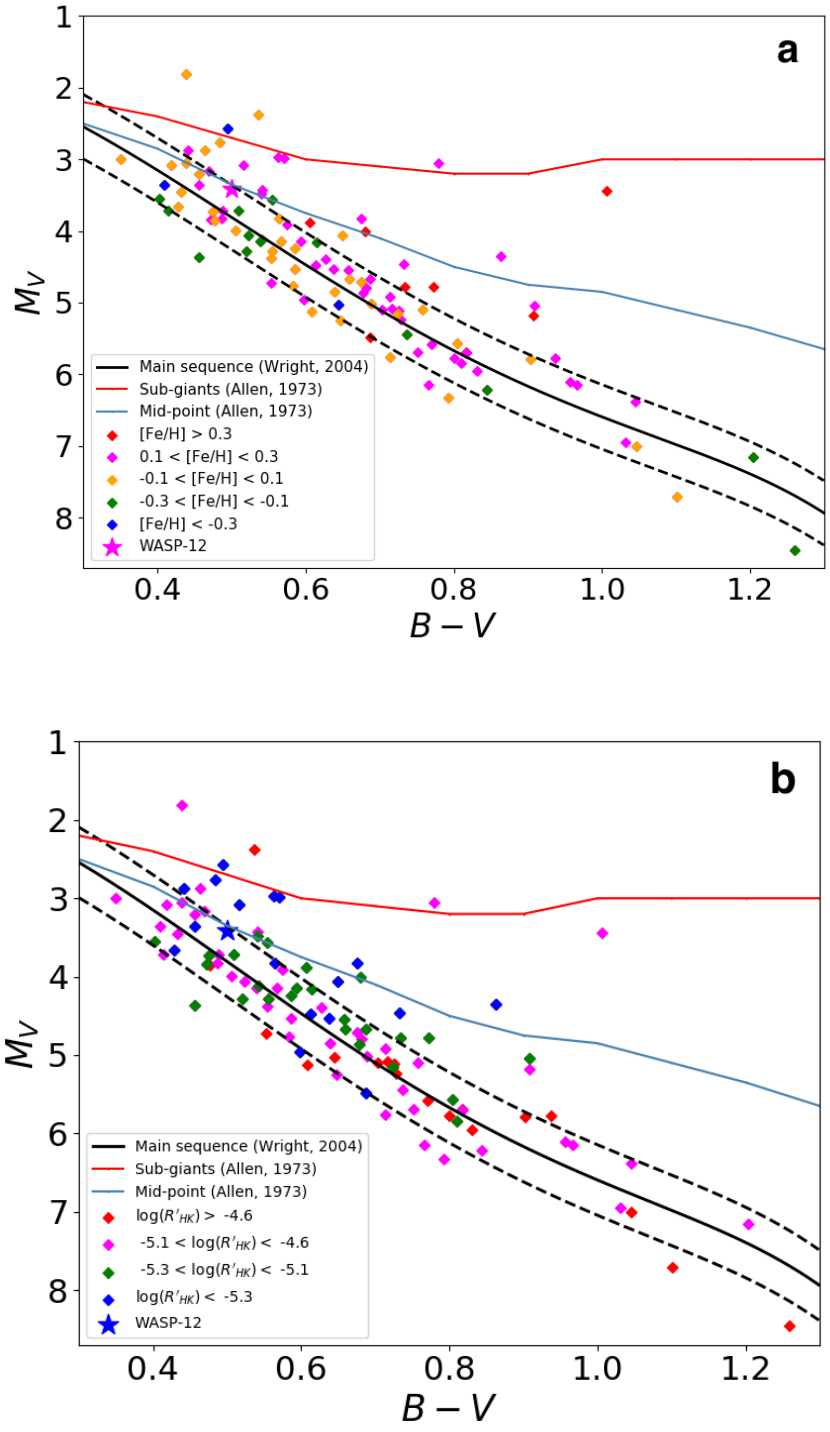

FIGURE 4.2: Colour magnitude diagrams for the OU-SALT sample with (a) metallicity and (b) activity colour-coded categories. The solid black line is the empirical average MS from Wright (2005b). The dashed black lines show a 0.45 mag range, corresponding to unevolved stars with $-0.3 < [\mathrm{Fe/H}] > +0.3$ (Wright et al., 2004). The sub-giant branch from Allen (1973) is shown in red, with the blue line marking the mid-point between Allen's MS and SG branch. The well-studied hot Jupiter host WASP-12 is plotted as a star.

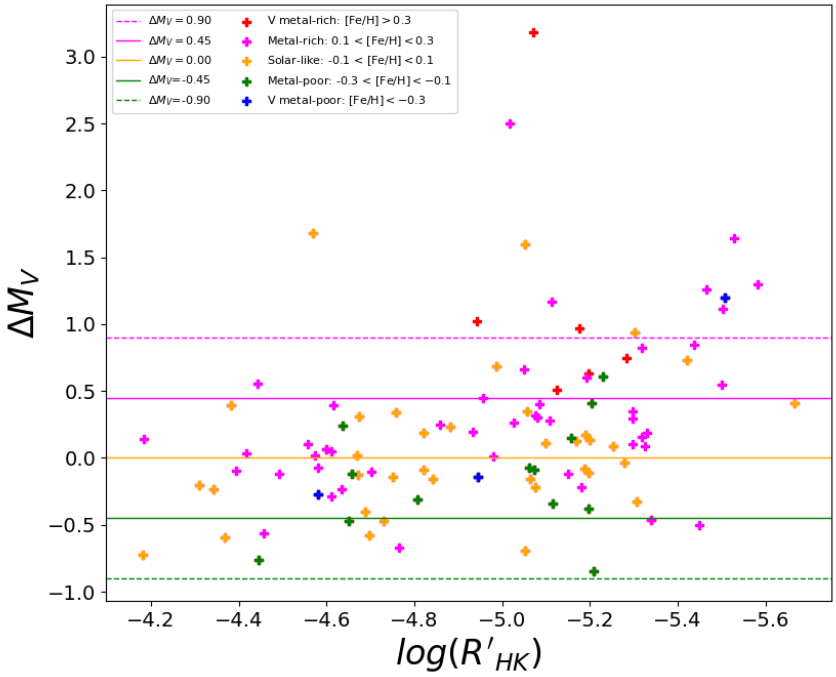

FIGURE 4.3: Activity plotted against ΔM_V, being the vertical difference between stellar magnitude and the empirical main sequence of Wright (2005b) shown an orange horizontal line. Marker colours indicate metallicity categories adapted from Lubin et al. (2012). The solid pink and green lines are 0.45 mag shifts from the MS.

4.2.3 Host classification

Several different methods of classifying evolved stars have been used. Wright (2004) adopts $\Delta M_V > 1$ as a "conservative limit for identifying evolved sub giants that exclude all but the most metal-rich stars". Judge & Saar (2007) treat the mid-point between the MS and SG branches (blue line in Figure 4.2) as the threshold above which stars may be considered evolved. Staab et al. (2017) define an unevolved sample of stars less than 0.45 mag above the MS of Wright (2005b), noting that this is an "arbitrary choice representing a compromise between retaining metal-rich, unevolved stars and rejecting metal-poor, evolved stars". Boro Saikia et al. (2018) use the main sequence $(B-V) - M_V$ relation from Table B.1 of Gray et al. (2006) to remove stars lying outside the relation by more than $|M_V| \pm 1$.

Gaia DR2 isochrones

A preferable approach to classifying OU-SALT targets is to use Gaia DR2 magnitudes and isochrones. Gaia measures magnitude, colour and parallax with a single instrument, removing the possibility of systematic offsets between measurements taken by different instruments. Isochrones generated from DR2 data provide the most accurate stellar evolutionary tracks available. Absolute Gaia magnitude ($G_{\rm abs}$) may be calculated similarly to M_V.

To assess whether targets have evolved off the MS, and to address the magnitude-metallicity relationship described, the OU-SALT sample has been divided into five metallicity categories: [Fe/H] < -0.3; -0.3 < [Fe/H] < -0.1; -0.1 < [Fe/H] < 0.1; 0.1 < [Fe/H] < 0.3; and [Fe/H] > 0.3. Isochrones corresponding to the mid-metallicity of each category are generated from the Dartmouth Stellar Evolution Database (Dotter et al., 2008) and Padova Isochrone Database (Girardi et al., 2004, 2008), and plotted with the targets on a HR diagram for each category (Figure 4.4). For example, for the -0.1 < [Fe/H] < 0.1 category, isochrones for stars with [Fe/H] = 0 are plotted. The Padova isochrones (which use the [M/H] metallicity index) are plotted in red and the Dartmouth ([Fe/H] index) in grey. The terminal age of the main sequence (TAMS), identified from the 'label' column in the Padova isochrones, is plotted as a solid black line. Uncertainties in $G_{\rm abs}$, propagated from Gaia DR2 distances, are often smaller than the plotted markers. $T_{\rm eff}$ uncertainties are considerably larger.

Systems above the TAMS are classified as evolved while those below are MS stars. In the [Fe/H] < -0.3 bin, WASP-87A—the lowest metallicity target in the OU-SALT sample with [Fe/H] $= -0.41$—is proximate to the TAMS, indicating it is approaching the end of its MS life. In the -0.3 <[Fe/H]< -0.1 bin, there are two high magnitude stars, K2-135 and WASP-80, that sit above the TAMS. However, due to the contours of the isochrones, the TAMS counter-intuitively passes beneath the MS in this region. Using the TAMS in Padova isochrones is therefore not advisable for classifying lower magnitude stars. However, both stars sit above the Dartmouth isochrones MS, with WASP-80 within the Padova MS. Given their proximity to the MS, and possible magnitude shifts due to metallicity, these targets are classified as

MS stars. WASP-127 is proximate to the TAMS and, indeed, Lam et al. (2017) confirms that the star is approaching the end of its MS life.

In the solar-like bin (-0.1 <[Fe/H]< -0.1), classifications are straightforward except in the case of WASP-103. If the Gaia DR2 distance for WASP-103 ($d = 833^{+178.43}_{-127.76}$ pc) is used to calculate G_{abs}, the star is a clear SG. However, as discussed above, the uncertainties on the WASP-103 Gaia DR2 distance are large. If the Hipparcos distance value ($d = 470 \pm 35$ pc) is used then the system is classified as a MS target. Gillon et al. (2014) note that spectroscopy and time-series radial velocity measurements have an amplitude consistent with a planet orbiting a MS star. Maciejewski et al. (2018) state tidal quality parameters confirm WASP-103 is on the MS. Thus, WASP-103 is assumed to be a MS host. The Padova tool could not generate isochrones for [M/H] = 0.4, so [M/H] = 0.3 is used instead in the [0.1 < [Fe/H] < 0.3] bin. WASP-47 and WASP-104 plot close to the TAMS. However, if compared to the Darmouth isochrone, they sit more centrally in the MS and so retain their MS status.

Tables 4.1 and 4.2 detail stellar parameters for the MS and SG sub-samples respectively. There are 80 MS targets and 24 evolved SG hosts. Comparing this division to other approaches: the method adopted by Staab et al. (2017) provides approximately the same ratio of MS:SG targets, however a handfull of systems receive different classifications. The approach of Wright (2004) and Boro Saikia et al. (2018) are insufficiently stringent for identifying MS stars, with around half as many SGs identified. For example, WASP-114 (Barros, S. C. C. et al., 2016) is a MS target under the Wright (2004) approach but clearly evolved when using DR2 isochrones.

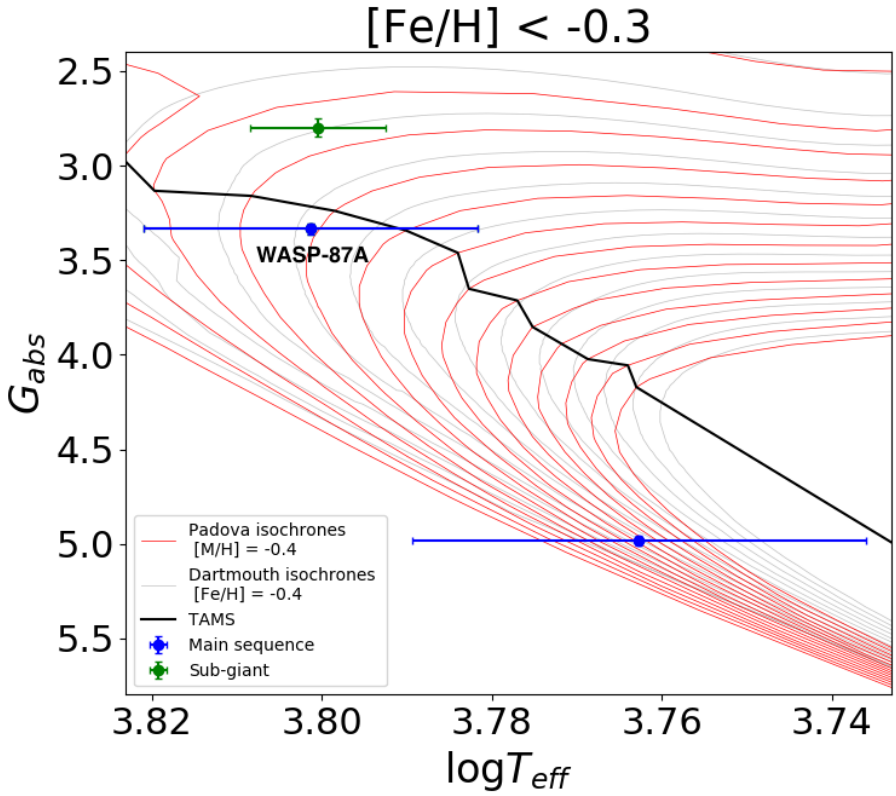

FIGURE 4.4: The logarithm of Gaia-measured effective temperature plotted against Gaia absolute magnitudes for OU-SALT targets. Dartmouth isochrones are plotted in silver for metallicities [Fe/H] = -0.4, [Fe/H] = -0.2, [Fe/H] = 0.0, [Fe/H] = 0.2 and [Fe/H] = 0.4 in separate panels. Corresponding Padova isochrones are plotted in red for metallicities [M/H] = -0.4, [M/H] = -0.2, [M/H] = 0.0, [M/H] = 0.2 and [M/H] = 0.3. The TAMS is shown as a solid black line. Where hosts have evolved above the TAMS, they are classified as SGs.

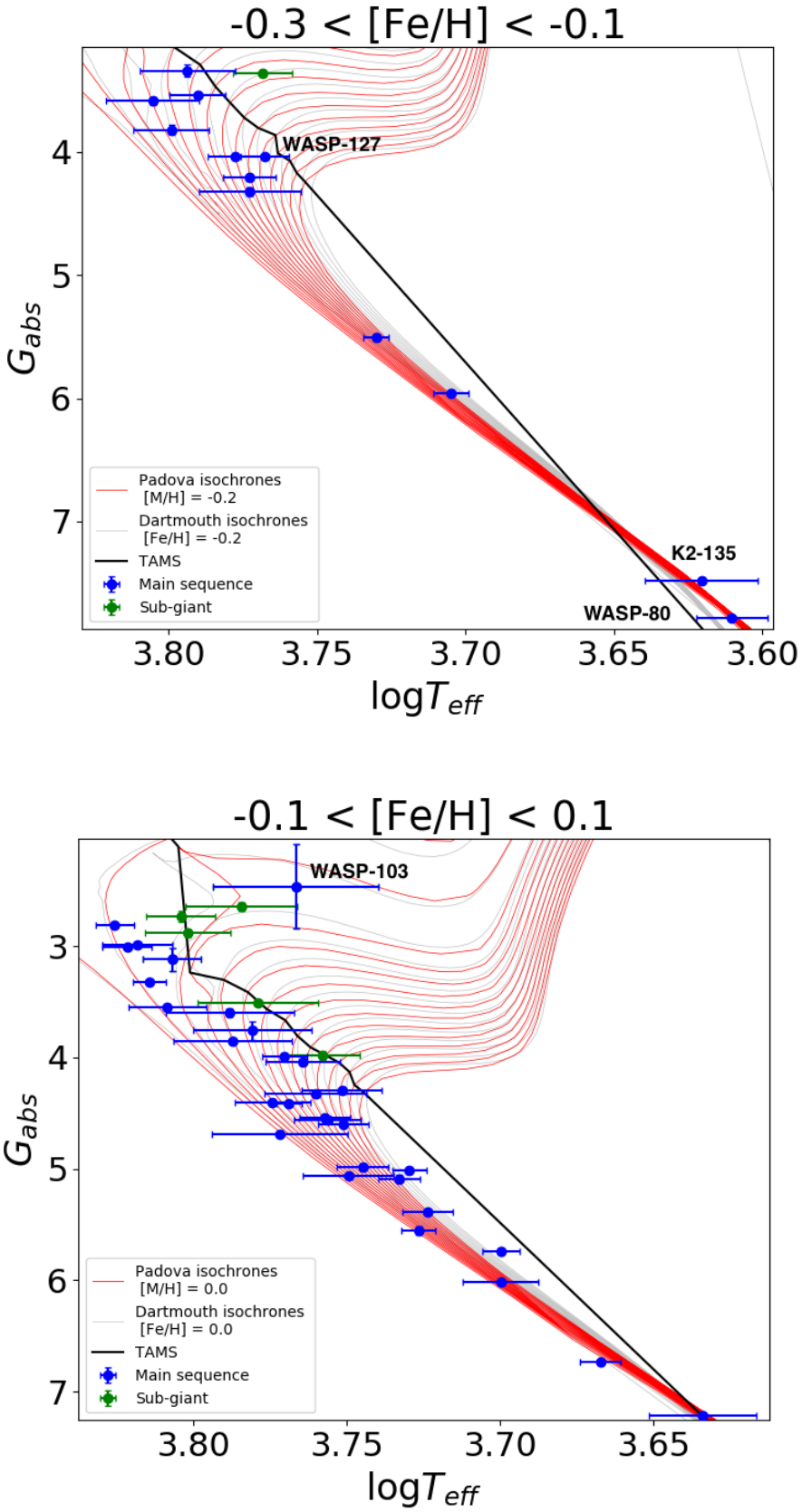

FIGURE 4.4: See caption above.

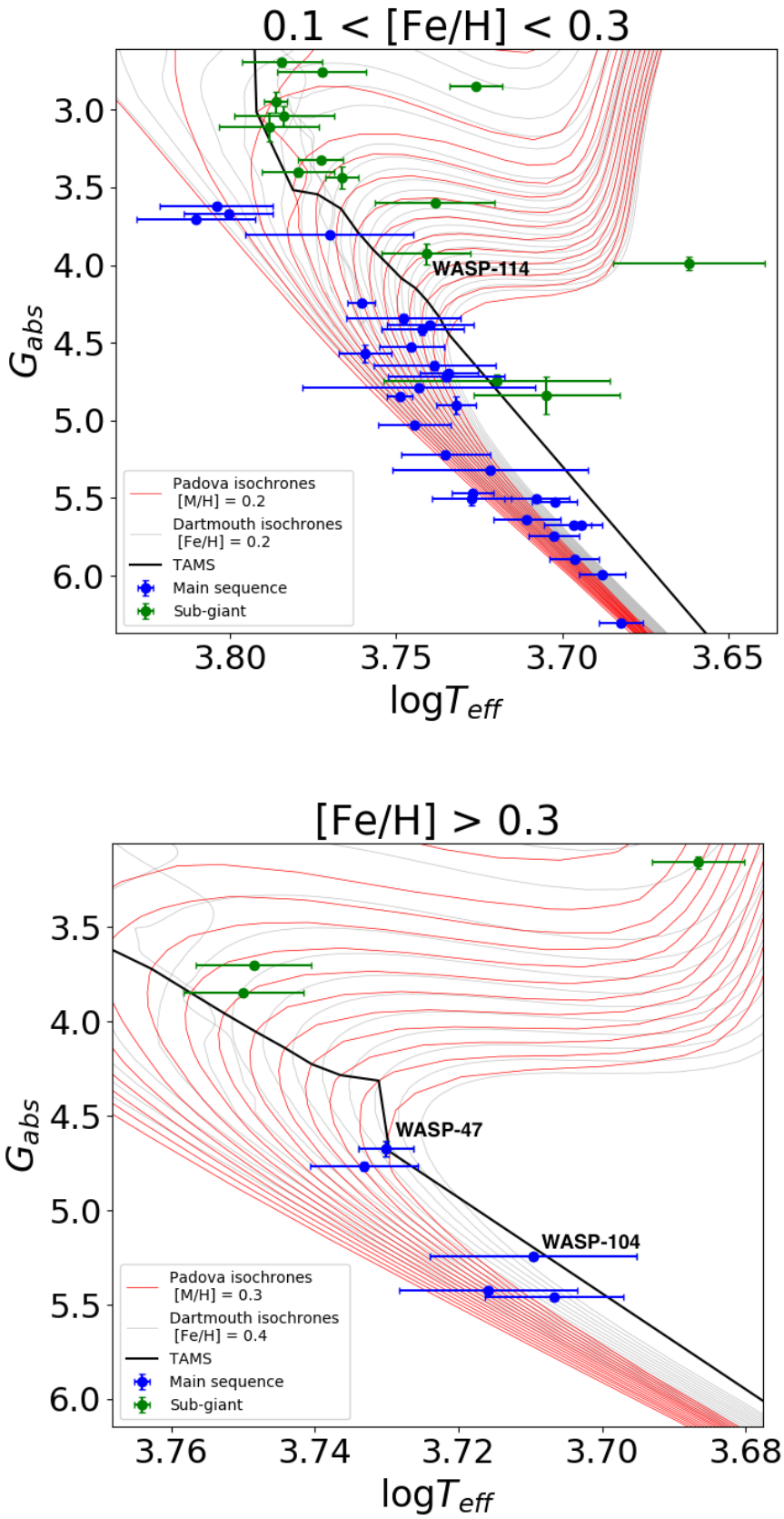

FIGURE 4.4: See caption above.

TABLE 4.1: Stellar parameters for OU-SALT MS targets: visual magnitude (V), absolute magnitude (M_V), distance (d), effective temperature (T_{eff}), stellar mass ($M_\star$), stellar radius ($R_\star$) and metallicity ([Fe/H]), and rotation period upper limit ($P_{rot}/\sin i$).

Target	V[a]	M_V[b]	d[c]	T_{eff}[d]	$M_\star$[d]	$R_\star$[d]	[Fe/H][d]	$P_{rot}/\sin i$[e]
			[pc]	[K]	[$M_\odot$]	[$R_\odot$]	[dex]	[days]
CoRoT-7	11.7	5.71	160±0.9	5259 ± 58	0.91±0.017	0.82±0.019	0.14 ± 0.06	23.0 (1)
CoRoT-11	12.9	3.86	660±24.9	6343 ± 72	1.26±0.140	1.37±0.061	0.04 ± 0.03	1.74 (2)
EPIC 219388192	12.5	5.12	310±4.6	5850 ± 50	1.01±0.040	1.01±0.030	0.03 ± 0.08	12.6 ± 2.1 (3)
HAT-P-27	12.2	5.69	200±1.5	5316 ± 55	0.94±0.035	0.90±0.054	0.30 ± 0.03	113.6 ± 113.8
HATS-2	13.6	5.95	340±3.7	5227 ± 95	0.88±0.037	0.90±0.019	0.15 ± 0.05	30.3 ± 10.1
HATS-3	12.4	4.36	410±7.9	6351 ± 76	1.21±0.036	1.40±0.030	-0.16 ± 0.07	7.8 ± 1.1
HATS-10	13.1	4.47	530±7.7	5880 ± 120	1.10±0.054	1.10±0.055	0.15 ± 0.10	9.8 ± 1.3
HATS-13	13.9	5.76	420±8.5	5523 ± 69	0.96±0.029	0.89±0.019	0.05 ± 0.06	15.9 ± 1.7
HATS-18	14.1	5.09	620±16.1	5600 ± 120	1.04±0.047	1.02±0.057	0.28 ± 0.08	9.8 ± 0.4
HATS-21	12.8	5.49	290±3.0	5695 ± 67	1.08±0.026	1.02±0.089	0.30 ± 0.04	19.6 ± 4.5
HATS-27	12.8	3.21	820±41.1	6438 ± 64	1.42±0.048	1.74±0.170	0.09 ± 0.04	9.4 ± 1.1
HATS-29	12.6	4.87	350±5.0	5670 ± 110	1.03±0.049	1.07±0.038	0.16 ± 0.08	23.1 ± 7.9
HATS-30	12.2	4.53	340±2.9	5943 ± 70	1.09±0.031	1.06±0.039	0.06 ± 0.05	13.1 ± 1.7
HATS-36	14.4	4.73	860±21.9	6149 ± 76	1.22±0.029	1.16±0.039	0.28 ± 0.04	14.0 ± 0.8
HD 73256	8.1	5.24	37±0.1	5570 ± 50	1.05	0.89	0.29 ± 0.05	13.97 (4)
K2-2	10.2	6.21	62±0.2	5089 ± 50	0.78±0.027	0.72±0.024	-0.16 ± 0.08	16.0 (5)
K2-19	13.0	5.69	290±5.7	5430 ± 60	0.93±0.050	0.86±0.040	0.10 ± 0.04	20.5 ± 0.3
K2-31	10.8	5.58	110±0.7	5412 ± 34	1.00±0.064	0.99±0.070	0.20 ± 0.03	18.4 ± 0.3
K2-32	12.3	6.32	160±1.3	5275 ± 60	0.86±0.028	0.84±0.044	-0.02 ± 0.04	61.1
K2-96	8.9	5.57	47±0.1	5261 ± 60	0.86±0.040	0.86±0.040	0.04 ± 0.05	25.0 ± 15.0
K2-106	12.1	5.16	240±3.0	5496 ± 46	0.92±0.030	0.95±0.050	0.06 ± 0.03	17.2 ± 2.3
K2-135	10.4	7.16	44±0.1	4219 ± 70	0.65±0.060	0.64±0.063	-0.29 ± 0.12	16.9 ± 2.1
KELT-10	10.6	4.24	190±2.2	5948 ± 74	1.11±0.055	1.21±0.047	0.09 ± 0.11	17.0 ± 0.0
Qatar-2	13.3	7.01	180±1.2	4645 ± 50	0.73±0.024	0.70±0.008	-0.02 ± 0.08	10.9 ± 0.2
WASP-5	12.3	4.85	310±3.4	5770 ± 65	1.03±0.049	1.09±0.040	0.09 ± 0.09	18.1 ± 2.5
WASP-6	11.9	5.44	200±1.6	5375 ± 65	0.84±0.067	0.86±0.025	-0.20 ± 0.09	27.3 ± 4.7
WASP-7	9.5	3.45	160±1.3	6520 ± 70	1.32±0.072	1.48±0.088	0.00 ± 0.10	4.4 ± 0.6
WASP-8	9.9	5.10	90±0.4	5600 ± 80	1.03±0.054	0.94±0.051	0.17 ± 0.07	30.1 ± 2.4
WASP-15	10.9	3.66	280±3.4	6573 ± 70	1.30±0.051	1.52±0.044	0.09 ± 0.04	19.3 ± 9.6
WASP-17	11.6	3.55	410±8.8	6550 ± 100	1.29±0.079	1.58±0.041	-0.25 ± 0.09	8.9 ± 1.5
WASP-18	9.3	3.84	120±0.4	6400 ± 70	1.29±0.059	1.25±0.028	0.10 ± 0.08	5.8 ± 0.8
WASP-25	11.9	5.24	210±1.5	5736 ± 35	1.05±0.038	0.92±0.018	0.06 ± 0.03	15.6 ± 5.2
WASP-28	12.0	4.29	350±5.7	6084 ± 45	0.99±0.067	1.08±0.025	-0.20 ± 0.07	16.9 ± 1.8
WASP-32	11.3	4.06	280±3.3	6100 ± 100	1.10±0.030	1.11±0.050	-0.13 ± 0.10	11.7 ± 2.0
WASP-34	10.3	4.68	130±0.9	5704 ± 26	1.01±0.070	0.93±0.120	0.08 ± 0.04	33.6 ± 15.0
WASP-36	12.7	4.77	390±5.3	5928 ± 59	1.08±0.034	0.98±0.014	-0.01 ± 0.05	15.1 ± 5.5
WASP-38	9.4	3.72	140±0.8	6150 ± 80	1.20±0.036	1.33±0.030	-0.12 ± 0.07	7.8 ± 0.4
WASP-43	12.4	7.71	87±0.3	4520 ± 120	0.72±0.025	0.67±0.011	-0.01 ± 0.12	15.6 ± 0.4

Target	V	M_V	d	$T_{\rm eff}$	$M_\star$	$R_\star$	[Fe/H]	$P_{\rm rot}$ / sin i[a]
			[pc]	[K]	[$M_\odot$]	[$R_\odot$]	[dex]	[days]
WASP-44	12.9	5.09	360±5.7	5400 ± 150	0.92±0.092	0.86±0.038	0.06 ± 0.10	13.7 ± 3.9
WASP-45	11.8	5.18	210±1.8	5100 ± 200	0.90±0.067	0.92±0.024	0.36 ± 0.12	20.2 ± 6.2
WASP-46	12.9	5.03	380±4.5	5600 ± 150	0.83±0.076	0.86±0.027	-0.37 ± 0.13	16.0 ± 1.0
WASP-47	11.9	4.79	260±4.9	5576 ± 67	1.04±0.031	1.14±0.013	0.36 ± 0.05	32.0 ± 4.3
WASP-50	11.4	5.11	180±1.9	5518 ± 42	0.86±0.057	0.85±0.019	0.13 ± 0.04	16.6 ± 3.2
WASP-51	10.3	3.72	210±2.3	6338 ± 42	1.24±0.041	1.22±0.051	0.12 ± 0.03	28.0 ± 6.5
WASP-52	12.0	5.79	170±1.3	5000 ± 100	0.80±0.050	0.79±0.016	0.03 ± 0.12	11.8 ± 3.3
WASP-55	11.8	4.38	300±4.0	6070 ± 53	1.16±0.036	1.10±0.021	0.09 ± 0.04	18.0 ± 5.8
WASP-62	10.2	3.99	180±0.6	6230 ± 80	1.25±0.050	1.28±0.050	0.04 ± 0.06	7.4 ± 0.4
WASP-65	11.9	4.71	270±2.7	5600 ± 100	0.93±0.140	1.01±0.050	-0.07 ± 0.07	14.2 ± 2.1
WASP-67	12.5	6.15	190±1.6	5417 ± 85	0.83±0.062	0.82±0.022	0.18 ± 0.06	19.7 ± 3.8
WASP-69	9.9	6.38	50±0.1	4700 ± 50	0.83±0.029	0.81±0.028	0.15 ± 0.08	23.1 ± 0.2
WASP-70	10.8	4.05	220±2.9	5700 ± 80	1.11±0.042	1.22±0.089	-0.01 ± 0.06	34.2 ± 8.0
WASP-75	11.5	4.12	290±4.4	6100 ± 100	1.16±0.030	1.27±0.020	0.07 ± 0.09	14.9 ± 2.8
WASP-77	10.1	5.01	110±1.2	5605 ± 41	1.00±0.045	0.95±0.015	0.07 ± 0.03	15.4 ± 0.4
WASP-79	10.0	3.08	250±1.8	6600 ± 100	1.39±0.080	1.51±0.040	0.03 ± 0.10	4.0 ± 0.2
WASP-80	11.9	8.45	50±0.1	4145 ± 100	0.60±0.035	0.59±0.012	-0.14 ± 0.16	23.6 ± 3.2
WASP-87A	10.7	3.36	300±3.6	6450 ± 120	1.20±0.093	1.63±0.062	-0.41 ± 0.10	8.6 ± 0.7
WASP-89	13.1	5.78	290±2.7	4955 ± 100	0.92±0.080	0.88±0.030	0.15 ± 0.14	20.2 ± 4.0
WASP-91	12.0	6.10	150±0.5	4920 ± 80	0.84±0.070	0.86±0.030	0.19 ± 0.13	18.1 ± 3.1
WASP-94B	10.5	3.90	210±2.1	6040 ± 90	1.24±0.090	1.35±0.120	0.23 ± 0.14	21.7 (6)
WASP-95	10.1	4.40	140±0.8	5830 ± 140	1.11±0.090	1.13±0.080	0.14 ± 0.16	19.7 ± 3.9 (7)
WASP-96	12.2	4.46	350±4.7	5500 ± 150	1.06±0.090	1.05±0.050	0.14 ± 0.19	35.4 ± 30.8
WASP-97	10.6	4.67	150±0.5	5670 ± 110	1.12±0.050	1.06±0.040	0.23 ± 0.11	48.8 ± 22.2
WASP-100	10.8	2.99	360±2.7	6900 ± 120	1.57±0.100	2.00±0.300	-0.03 ± 0.10	7.9 ± 1.3
WASP-101	10.3	3.82	200±1.2	6380 ± 120	1.34±0.070	1.29±0.040	0.20 ± 0.12	5.3 ± 0.3
WASP-103	12.1	2.37	880±178.4	6110 ± 160	1.21±0.096	1.41±0.048	0.06 ± 0.13	6.7 ± 0.6
WASP-104	11.1	4.77	190±1.5	5450 ± 130	1.01±0.050	0.94±0.016	0.32 ± 0.09	118.9 ± 208.2
WASP-108	11.2	4.15	260±3.3	6000 ± 140	1.17±0.092	1.22±0.040	0.05 ± 0.11	12.8 ± 2.2
WASP-109	11.5	3.71	360±5.0	6520 ± 140	1.20±0.090	1.35±0.044	-0.22 ± 0.08	4.4 ± 0.1
WASP-117	10.1	4.15	160±0.6	6040 ± 90	1.13±0.029	1.17±0.067	-0.11 ± 0.14	17.1 ± 2.6
WASP-118	11.0	3.16	380±11.2	6410 ± 125	1.32±0.035	1.75±0.016	0.16 ± 0.11	9.2 ± 1.1
WASP-119	12.2	4.80	300±2.0	5650 ± 100	1.02±0.060	1.20±0.100	0.14 ± 0.10	86.8 ± 111.8
WASP-120	11.0	3.05	380±3.2	6450 ± 120	1.39±0.057	1.87±0.110	-0.05 ± 0.07	6.3 ± 0.6
WASP-123	11.0	4.55	200±3.1	5740 ± 130	1.17±0.061	1.28±0.051	0.18 ± 0.08	65.0 ± 45.6
WASP-127	10.2	4.16	160±1.2	5750 ± 100	1.08±0.030	1.39±0.030	-0.18 ± 0.06	234.5 ± 156.4
WASP-126	11.2	4.53	220±0.9	5800 ± 100	1.12±0.060	1.27±0.100	0.17 ± 0.08	128.6 ± 129.0
WASP-130	11.1	4.92	170±1.4	5600 ± 100	1.04±0.040	0.96±0.030	0.26 ± 0.10	97.2 ± 97.2
WASP-132	12.4	6.95	120±0.6	4750 ± 100	0.80±0.040	0.74±0.020	0.22 ± 0.13	33.0 ± 3.0
WASP-138	11.8	3.72	410±7.6	6300 ± 100	1.22±0.050	1.36±0.050	-0.09 ± 0.10	8.9 ± 1.3
WASP-139	12.5	5.84	210±1.1	5300 ± 100	0.92±0.100	0.80±0.040	0.20 ± 0.09	9.6 ± 2.6

Target	V	M_V	d	$T_{\rm eff}$	$M_\star$	$R_\star$	[Fe/H]	$P_{\rm rot}/\sin i$[a]
			[pc]	[K]	[$M_\odot$]	[$R_\odot$]	[dex]	[days]
WASP-140	11.1	5.77	120±0.6	5300 ± 100	0.90±0.040	0.87±0.040	0.12 ± 0.10	10.4 ± 0.1
WASP-156	11.6	6.15	120±0.6	4910 ± 61	0.84±0.052	0.76±0.030	0.24 ± 0.12	10.1 ± 2.5

[a] Values from SIMBAD.

[b] Calculated with Equation 3.17 of Green & Jones (2015).

[c] Values from BJ18.

[d] Values from TEPCat.

[e] Calculated using $v \sin i$ and $R_\star$ unless a reference is provided. References for published rotation periods: (1) Léger et al. (2009); (2) Gandolfi et al. (2010); (3) Nowak et al. (2017) (4) Udry et al. (2003); (5) Vanderburg et al. (2015b); (6) Neveu-VanMalle et al. (2014); (7) Hellier et al. (2014).

TABLE 4.2: Stellar parameters of OU-SALT SG targets. Symbols and footnotes as described in Table 4.1, unless otherwise stated.

Target	V[a]	M_V[b]	d[c]	$T_{\rm eff}$[d]	$M_\star$[d]	$R_\star$[d]	[Fe/H][d]	$P_{\rm rot}/\sin i$[e]
			[pc]	[K]	[$M_\odot$]	[$R_\odot$]	[dex]	[days]
CoRoT-22	13.9	4.96	630±11.2	5939 ± 120	1.10 ± 0.049	1.14 ± 0.090	0.17 ± 0.12	16.0 (1)
CoRoT-28	13.8	4.35	770±14.7	5150 ± 100	1.01 ± 0.140	1.78 ± 0.110	0.15 ± 0.10	30.0 (2)
K2-39	10.9	3.43	310±4.7	4912 ± 60	1.19 ± 0.085	2.93 ± 0.210	0.43 ± 0.04	12.1 (3)
KELT-11	8.0	3.06	99±0.5	5375 ± 25	1.44 ± 0.070	2.69 ± 0.040	0.17 ± 0.07	51.2 ± 9.7
KELT-15	11.4	3.83	320±2.5	6003 ± 56	1.18 ± 0.051	1.48 ± 0.091	0.05 ± 0.03	9.9 ± 0.8
WASP-26	11.3	4.29	250±4.8	6015 ± 55	1.09 ± 0.046	1.28 ± 0.036	-0.02 ± 0.09	16.7 ± 1.8
WASP-63	11.2	3.83	290±2.0	5715 ± 60	1.32 ± 0.050	1.88 ± 0.100	0.28 ± 0.05	34.0 ± 6.3
WASP-72	11.0	2.77	430±8.5	6250 ± 100	1.39 ± 0.055	1.98 ± 0.240	-0.06 ± 0.09	16.7 ± 2.8
WASP-73	10.5	2.97	320±3.0	6030 ± 120	1.34 ± 0.050	2.07 ± 0.190	0.14 ± 0.14	17.2 ± 2.3
WASP-74	9.8	3.88	150±1.1	5990 ± 110	1.48 ± 0.120	1.64 ± 0.050	0.39 ± 0.13	20.2 ± 4.0
WASP-76	9.5	3.08	190±6.2	6250 ± 100	1.46 ± 0.070	1.73 ± 0.040	0.23 ± 0.10	26.5 ± 4.9
WASP-78	12.0	2.57	750±17.1	6100 ± 150	1.39 ± 0.090	2.35 ± 0.100	-0.35 ± 0.14	16.5 ± 2.0
WASP-82	10.1	2.87	280±3.1	6500 ± 80	1.64 ± 0.079	2.22 ± 0.097	0.12 ± 0.11	43.2 ± 15.1
WASP-88	10.4	1.80	520±8.8	6430 ± 130	1.45 ± 0.050	2.08 ± 0.120	-0.08 ± 0.12	12.5 ± 1.4
WASP-90	11.7	3.36	470±19.9	6440 ± 130	1.55 ± 0.100	1.98 ± 0.090	0.11 ± 0.14	16.7 ± 1.6
WASP-94A	10.1	3.43	210±2.5	6170 ± 80	1.45 ± 0.090	1.62 ± 0.050	0.26 ± 0.15	19.5 ± 2.4
WASP-99	9.5	3.48	160±0.8	6150 ± 100	1.48 ± 0.100	1.76 ± 0.110	0.21 ± 0.15	13.1 ± 1.3
WASP-105	12.0	5.04	250±14.8	5070 ± 130	0.89 ± 0.090	0.90 ± 0.030	0.28 ± 0.16	26.8 ± 30.0
WASP-111	10.2	2.88	300±6.8	6400 ± 150	1.50 ± 0.110	1.85 ± 0.100	0.08 ± 0.08	8.4 ± 0.7
WASP-114	12.7	4.15	520±17.3	5940 ± 140	1.29 ± 0.053	1.43 ± 0.060	0.14 ± 0.07	11.3 ± 1.3
WASP-118	11.0	3.16	380±11.2	6410 ± 125	1.32 ± 0.035	1.75 ± 0.016	0.16 ± 0.11	9.2 ± 1.1
WASP-122	11.0	4.01	250±1.5	5720 ± 130	1.24 ± 0.039	1.52 ± 0.030	0.32 ± 0.09	23.3 ± 5.7
WASP-131	10.1	3.56	200±2.7	5950 ± 100	1.06 ± 0.060	1.53 ± 0.050	-0.18 ± 0.08	25.8 ± 7.8
WASP-142	12.3	2.98	730±24.2	6060 ± 150	1.33 ± 0.080	1.64 ± 0.080	0.26 ± 0.12	26.8 ± 12.2

[e] Calculated using $v \sin i$ and $R_\star$ unless a reference is provided. References for published rotation periods: (1) Moutou et al. (2014); (2) Cabrera et al. (2015); (3) Van Eylen et al. (2016).

4.3 Stellar age

The relationships between age, activity and rotation are well-defined. Mamajek & Hillenbrand (2008) characterize activity age (τ_{act}) using a fit that is appropriate between $-4.0 < \log(R'_{HK}) < -5.1$ and for $6.7 < \log(\tau) < 9.9$. The ages of OU-SALT targets calculated using this method are shown in Table 3.1. Stellar rotation may also be used to estimate age. Gyrochronology derives ages for low-mass MS stars using rotation periods. Equation 3 of Barnes (2007) provides a widely cited expression for calculating stellar age. Gyro ages for the OU-SALT population are also detailed in Table 3.1. Figure 4.5 shows that there is a general agreement between gyro and chromospheric ages for a sample of Mount Wilson stars but that the gyro ages are roughly 25% lower than the chromospheric ages overall for bluer stars. There is good agreement for stars with of $B-V > 0.6$. No such agreement is present between the chromospheric and gyro ages of the OU-SALT population, as is evident from Figure 4.6. Activity ages stack up at $\sim$ 11 Gyr, while gyro ages are mainly below 8 Gyr. Counter intuitively, there is a group of low-activity targets with young gyro ages. There is also a group of three higher activity targets (HATS-2, WASP-6, WASP-50) with old gyro ages, possibly indicating strong enhancement of magnetic activity.

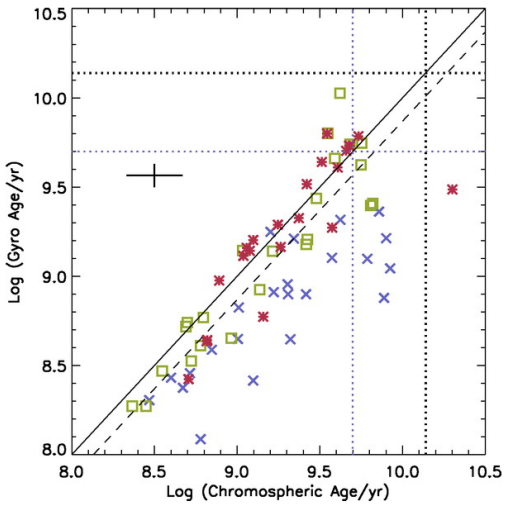

FIGURE 4.5: The gyro and chromo ages of the Mount Wilson stars compared. Crosses show stars bluer than $B-V = 0.6$, asterisks are stars redder than $B-V = 0.8$, and squares those with colours between. The solid line indicates equality, while the parallel dashed line ($age_{gyro} = 0.74\, age_{chromo}$) bisects the sample. (Barnes, 2007)

Barnes (2007) notes that chromospheric ages are problematic as emission varies during the rotation phase and the stellar cycle, and due to binarity Wilson (1963). As such, repeated measurements are required to obtain an average measurement of chromospheric emission. As it is difficult derive an accurate stellar age from a single measurement, chromospheric ages of the OU-SALT population should be treated cautiously. Indeed, (Staab et al., 2017) argue that activity metrics are an unreliable age indicator for close-in planet hosts as absorption (due to mass loss) can suppress, and SPI elevate, observed activity (Haswell et al., 2012; Cuntz et al., 2000). Similarly, Gallet & Delorme (2019) note that age estimation techniques such as gyrochronology and magnetochronoloy cannot be applied to stars whose rotational evolution has been influenced through the exchange of angular momentum by SPI. They found that gyrochronology can be applied to systems where the planetary orbital period is $P_{\rm orb} \gtrsim 4$ days, for a 1 $M_\odot$ star and a $M_{\rm P} < 5\,M_{\rm J}$ planet. Activity and gyro ages are thus not in general reliable in the OU-SALT population. Angus et al. (2019) advocate an approach to age-dating F, G, K and M field stars that combines gyrochronology and isochrone fitting. Asteroseismology may also be used to accurately estimate stellar age (Lebreton & Montalbán, 2008).

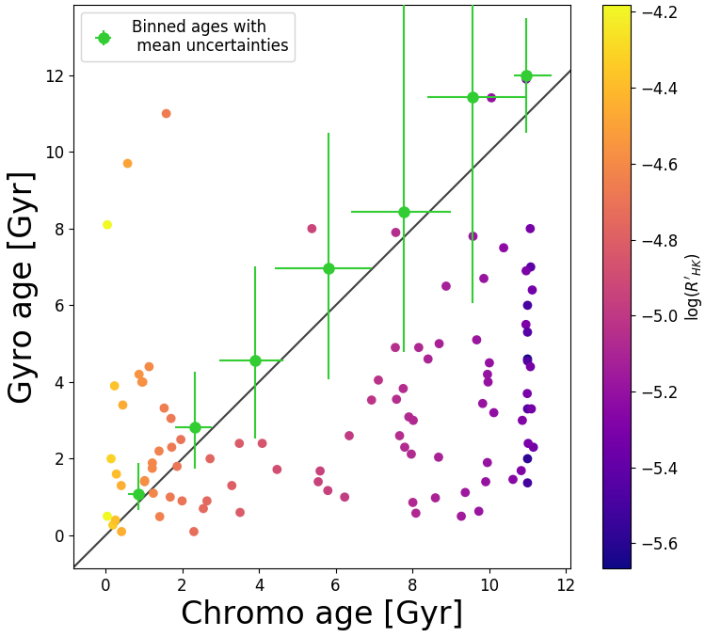

FIGURE 4.6: Comparison of the gyro and chromo ages of OU-SALT hosts calculated using Equation 3 from Barnes (2007) and Mamajek & Hillenbrand (2008). The colour scale shows activity. A diagonal line of equality is plotted in black. The data is binned and the mean age in each bin plotted with green markers. Mean uncertainties within each bin are indicated.

4.4 Planets

Planetary parameters for the OU-SALT sample are presented in Table 4.3.

4.4.1 Demographics

The planets of the OU-SALT sample comprise 4 super-Earths, 7 hot-Neptunes, 92 hot-Jupiters and a brown dwarf (Figure 4.7). The approximate position of the border of the Neptune Desert (ND: Mazeh et al., 2016) is shown in Figures 4.7(a) and 4.7(b). Hot Jupiters stack up on the upper border of the ND, with a handful of systems encroaching into the desert. Hot Neptunes are located near the apex of the ND border, while super Earths nestle along the lower border and within the desert.

Figure 4.7(a) shows the period-mass distribution with a colour map showing the logarithm of the surface gravity of the planet ($\log g_P$). Hot Jupiters near the upper boundary of the ND have the lowest g_P values, making these the systems most vulnerable to mass loss (Fossati et al., 2015a). Low mass super-Earths have lower g_P than some of the HJ population, also making these likely candidates for atmospheric escape. Within the HJ population, there is a trend of increasing g_P with mass, which is repeated on smaller scales in both the sub-Neptune and super-Earth populations. Figure 4.7(d) shows the radius distribution of planets. The ND is apparent in the plot, and the super Earth population straddles the Evaporation Valley (EV: Fulton et al., 2017; Van Eylen et al., 2018).

4.4.2 Incident flux

The flux incident upon exoplanets (S_P) may be estimated by the expression for radiation from a black body, and normalised to Earth units.

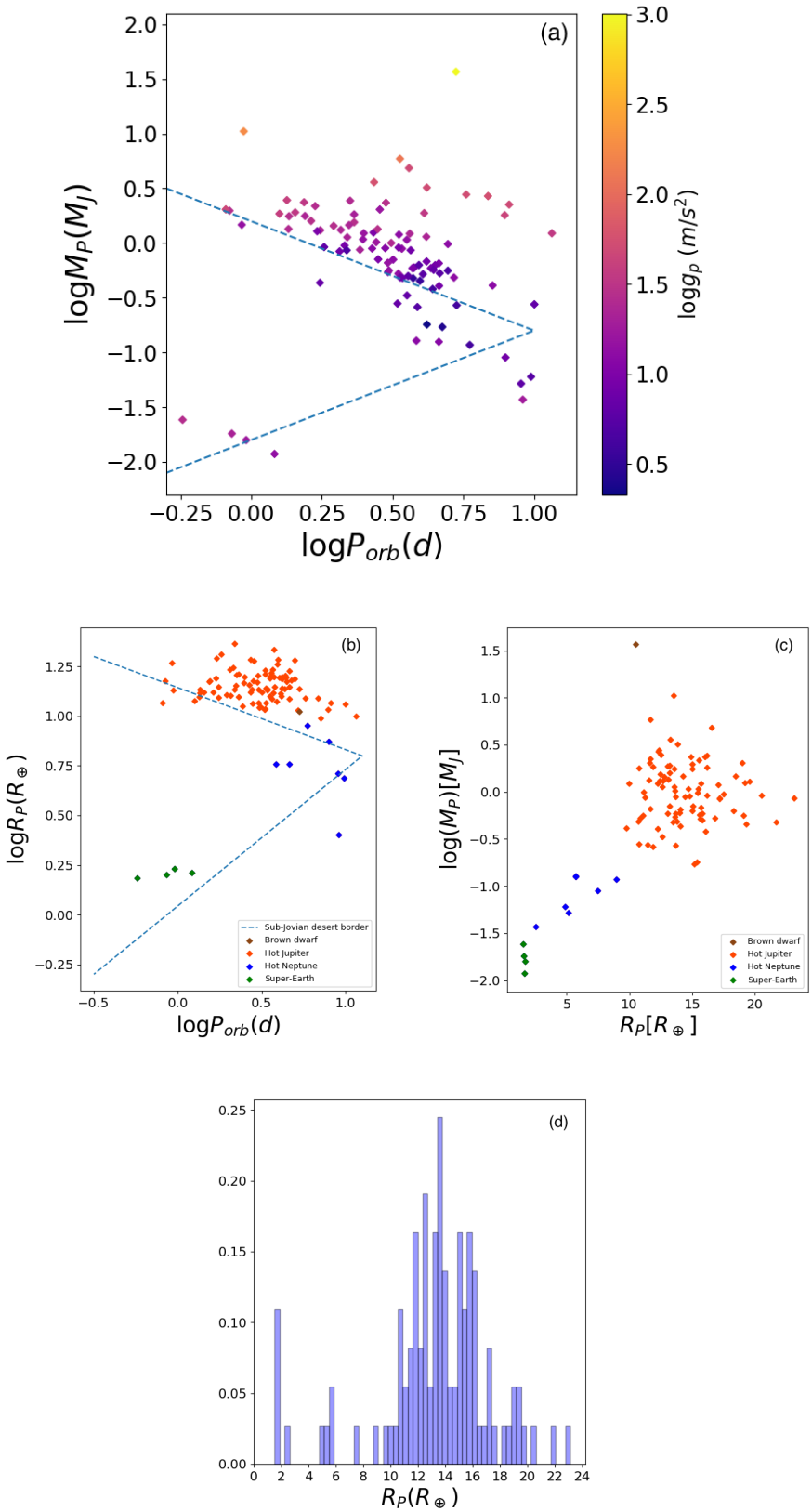

FIGURE 4.7: OU-SALT population demographics. Sample contains 1 brown dwarf, 93 hot-Jupiters, 7 hot-Neptunes and 4 super-Earths. Plots shows: (a) Period-mass distribution with a colourmap showing $\log g_P$; (b) Period-radius distribution; (c) Radius-mass distribution; and (d) Radii histogram with SJD and Evaporation Valley evident. The borders of the ND (from Mazeh et al., 2016) are shown by dashed lines in (a) and (b).

This provides:

$$S_{\rm P} = \frac{R_\star^2 (T_{\rm eff}/5777)^4}{a^2}, \tag{4.1}$$

where $R_\star$ is the radius of the star in solar radii ($R_\odot$), $T_{\rm eff}$ is the stellar effective temperature in K, and a is semi-major axis in AU (see Equation 9 of Coughlin et al., 2015, though note I have substituted $R_\star$ for $R_{\rm P}$).

Estimated incident flux values for OU-SALT systems are presented in Table 4.3. OU-SALT planets receive between 32 (K2-2) and 6379 (WASP-103) the level of incident flux received by the Earth, as shown in Figure 4.8. K2-106, WASP-18, WASP-72, WASP-76, WASP-78, WASP-82, WASP-87 A, WASP-100, WASP-103 and WASP-111 are notable for receiving more than 3000 times the flux of Earth. Intuitively, planets with the smallest separation distances receive the highest levels of incident flux. High irradiation levels (particularly at X-ray and far UV wavelengths) make mass loss by hydrodynamic escape more likely (Kubyshkina et al., 2018a,b; Owen, 2019).

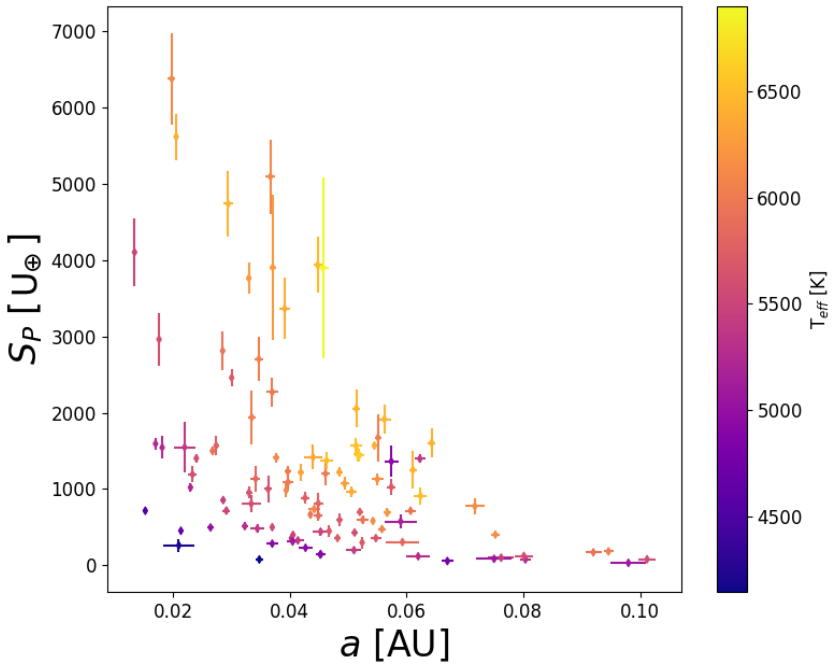

FIGURE 4.8: Semimajor axis plotted against incident flux in Earth units for the OU-SALT sample, with a colour scale showing effective temperature of the host. OU-SALT planets are exposed to between $32 - 6379$ (WASP-103) the level of incident flux received by the Earth.

4.4.3 Roche lobe filling

The Roche potential describes the gravitational equipotentials of close, co-rotating bodies. The 'Roche lobe' is the outer border of the region in which one body is gravitationally dominant (Paczynski, 1971). The lobes of the bodies meet at the inner Lagrangian (L_1) point. If one body's Roche lobe overfills, mass transfer will occur via the L_1 point. Using the Roche potential surface mapping approach of Busuttil (2017), we have generated Roche lobe filling factors for the OU-SALT population, assuming orbital inclination of 90°. The most distorted planet is WASP-103 (Gillon et al., 2014), with a filling factor of 54.12 % (Figure 4.9). Figure 4.10 shows the filling factors for the whole OU-SALT sample (except the two non-transiting targets). The next most distorted planets are HATS-18 (Penev et al., 2016), WASP-76 (West, R. G. et al., 2016), and WASP-78 (Smalley, B. et al., 2012), with volume filling percentages of 24.95 %, 29.02 % and 30.75 % respectively. Based on these calculations, mass loss through overfilling of the Roche-lobe due to planetary distortion is not in action in the OU-SALT population, even in the most extreme cases.

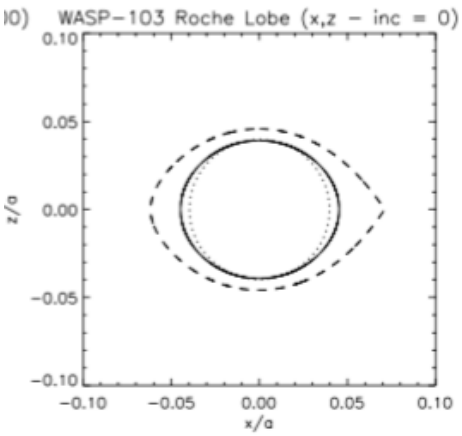

FIGURE 4.9: Cross sectional plot of WASP-103 in the $x-z$ plane, showing how the planetary surface deviates from spherical. The solid, dashed and dotted lines are mapped planetary surface, Roche lobe, and a perfect circle respectively. (Plot by R. Busuttil)

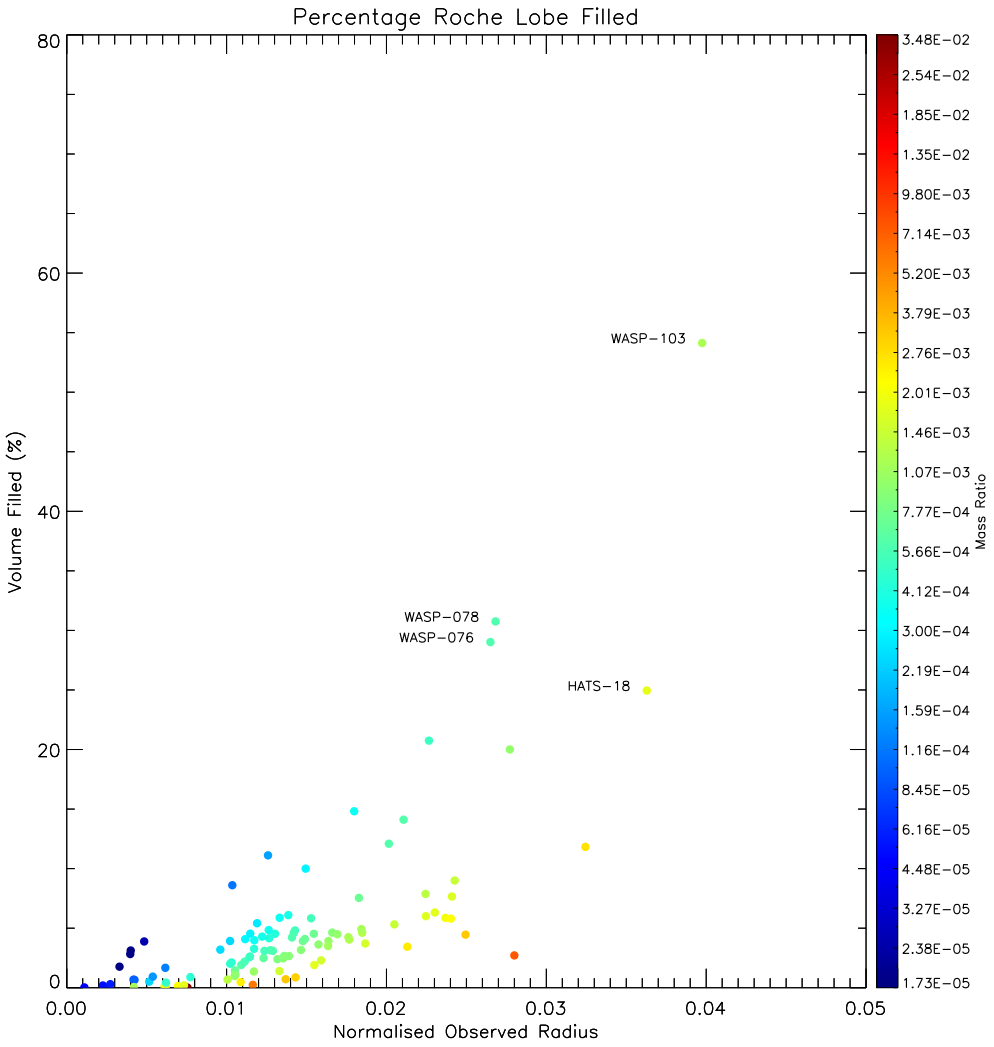

FIGURE 4.10: Roche lobe filling percentages as a function of normalised radius. The colour scale represents the planet-star mass ratio. The four planets with the largest filling factors—WASP-103, HATS-18, WASP-76 and WASP-78—are labelled. (Plot by R. Busuttil.)

4.5 Star-planet interaction proxies

Various expressions may used to estimate SPI strength.

4.5.1 Tidal star-planet interactions

A useful indicator of star-planet tidal interactions (SPTI) are estimates of angular momentum of the stellar rotation ($J_\star$) and planetary orbit ($J_{\rm orb}$):

$$J_\star = \frac{4}{5}\pi M_\star R_\star^2 P_{\rm rot}^{-1}, \tag{4.2}$$

and

$$J_{\rm orb} = 2\pi M_{\rm P} a^2 P_{\rm orb}^{-1}, \tag{4.3}$$

where a is the semi-major axis, $R_\star$ is the stellar radius, $M_\star$ and $M_{\rm P}$ are the masses of the star and planet respectively, $P_{\rm rot}$ is the stellar rotation period, and $P_{\rm orb}$ is planetary orbital period. Table 5.2 provides these values for each OU-SALT target. Where $J_{\rm orb} \gtrsim J_\star$ and the planet has migrated towards the host, sufficient angular momentum exists to inhibit stellar spin-down at the expected rate (Poppenhaeger & Wolk, 2014). Figure 4.11 shows that just under half of OU-SALT targets have $J_{\rm orb} \gtrsim J_\star$, providing an ideal sub-sample for the observation of SPTI.

Cuntz et al. (2000) use the tidal bulge height as a fraction of the scale height of the stellar photosphere ($h_{\rm tide}/h_{\rm scale}$) to calculate SPTI strength:

$$h_{\rm tide} = \frac{\Delta g_\star}{2g_\star} R_\star, \tag{4.4}$$

where $g_\star$ is the surface gravity of the star and $R_\star$ is the stellar radius. In addition:

$$\frac{\Delta g_\star}{g_\star} = \frac{M_P}{M_\star} \frac{2R_\star^3}{(a - R_\star)^3}, \tag{4.5}$$

and

$$h_{\text{scale}} = \frac{kT_{\text{eff}}}{m_H g_\star}, \tag{4.6}$$

where $M_\star$ and M_{P} are the masses of the star and planet respectively, m_H is the mean mass of an atom of Hydrogen, k is the Boltzmann constant, and T_{eff} is the effective temperature. Table 5.2 provides $h_{\text{tide}}/h_{\text{scale}}$ values for each target.

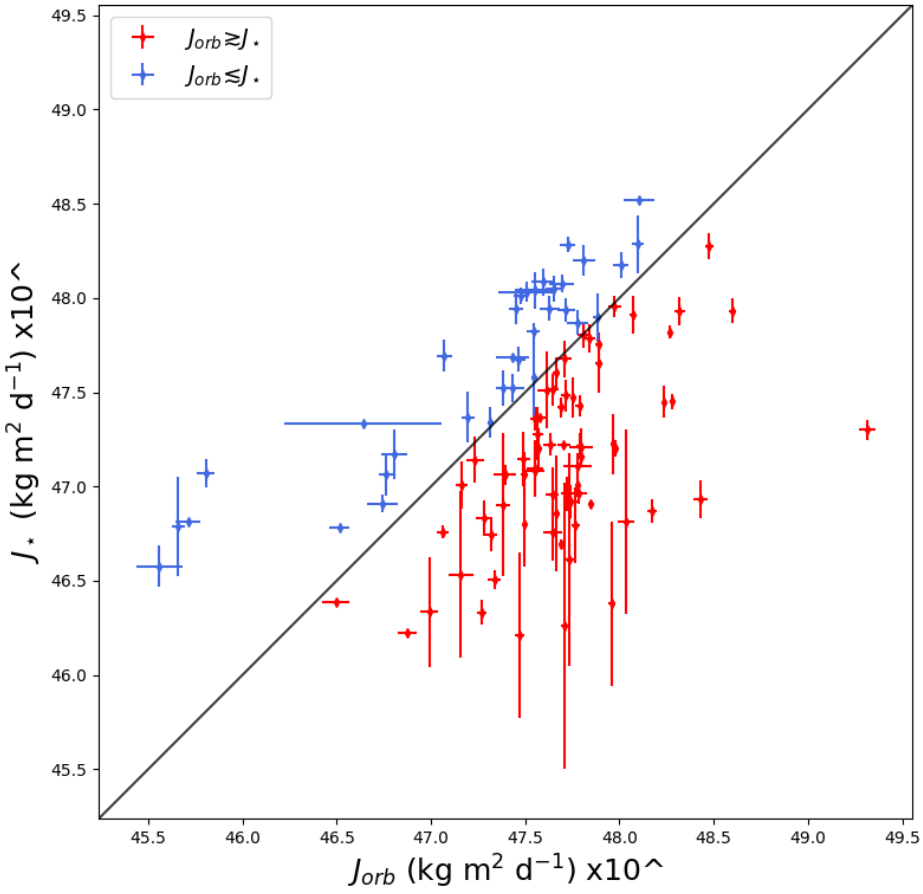

FIGURE 4.11: Stellar angular momentum plotted against angular momentum of the planetary orbit. Axes scales show powers. Points in red show systems where orbital angular momentum is similar to or greater than rotation angular momentum.

4.5.2 Magnetic star-planet interactions

The energy flux due to star-planet magnetic interactions (SPMI) scales with $B_\star B_P v_{\rm rel} a^{-n}$, where $B_\star$ and B_P are the mean magnetic field strengths averaged over the surface of the star and planet, $v_{\rm rel}$ is the relative velocity of the star and planet, and a is separation (Cuntz et al., 2000; Lanza, 2009; Iro & Deming, 2010; Lanza, 2012). The exponential n is ~ 2 in the open field region of the stellar wind, and ~ 3 when in the dipole-like field close to the star. While $B_\star$ can be estimated from stellar magnetic field measurements, it is more difficult to estimate B_P. Ideally, the power released by magnetic reconnection events should be calculated analytically (as per Lanza, 2013). However, calculation requires values of the stellar and planetary field strength, which have not been measured for OU-SALT targets. Instead simple proxies, such as $1/a$ and M_P/a^2 (Miller et al., 2015; Staab et al., 2017), may be used to estimate SPMI. Figure 4.12 shows logarithms of the simple SPMI proxy ($1/a$) plotted against the SPTI term ($h_{\rm tide}/h_{\rm scale}$). A positive monotonic relationship is apparent—the smaller the separation, the larger the strength of interaction between the bodies. Approximately half of the population are "extreme" systems with $M_P/a^2 > 450\ M_J$ AU^{-2} (Miller et al., 2015). This sub-sample provides an ideal test-bed for SPI.

4.6 Temporal activity variations

A dozen OU-SALT targets have been observed multiple times. The positions of these systems in the tidal-magnetic SPI plot (Figure 4.12) are indicated in Figure 4.13.

4.6.1 WASP-43

Despite having one the shortest known HJ orbital periods (0.81 d), WASP-43b receives relatively low levels of radiation (716 $U_\oplus$) from its K-type host ($T_{\rm eff} = 4520$ K Hellier et al., 2011). It has a 15.6 d rotation period and corresponding gyrochronological age of $300 - 600$ Myr—an estimate that is compatible with the star's X-ray profile (Czesla et al., 2013; Staab et al., 2017). These parameters may be effected by tidal spin-up. Zhao et al. (2018) find WASP-43b may have an orbital decay rate of

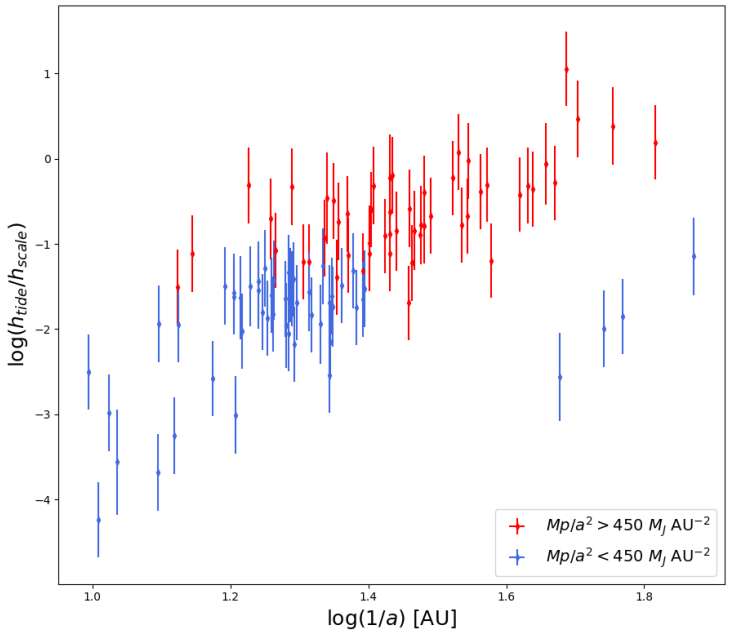

FIGURE 4.12: The inverse logarithm of semi-major axis (SPMI proxy) plotted against the logarithm of $h_{\rm tide}/h_{\rm scale}$ (SPTI term). The x-axis uncertainties are smaller than the markers. Approximately half of the OU-SALT population (plotted red) are "extreme systems", with $M_{\rm P}/a^2 > 450\,{\rm M_J\ AU^{-2}}$ (Miller et al., 2015).

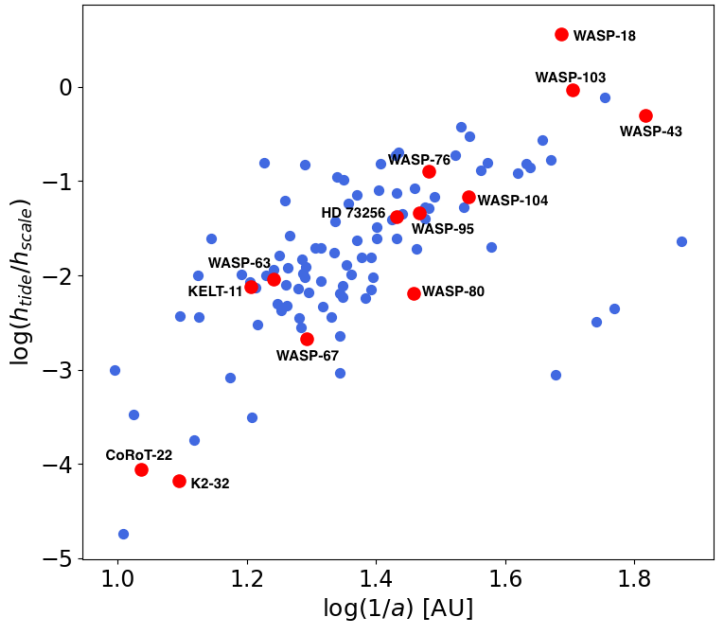

FIGURE 4.13: SPMI vs. SPTI proxies. Targets for which there are time series data are labelled. The position of the extreme interactor WASP-18 is also indicated.

-0.005248 ± 0.001714 s yr^{-1}. Esposito et al. (2017) report a spin-orbit angle of zero, and suggest SPI has been effective in circularising and aligning the planetary orbit. This could however also result from planetary formation processes. Staab et al. (2017) reported $\log(R'_{\rm HK}) = -4.17 \pm 0.10$ for WASP-43 at phase $\phi = 0.661$. Esposito et al. (2017) use HARPS-N spectroscopic data to find $\langle \log(R'_{\rm HK}) \rangle = -4.35 \pm 0.10$, which agrees with Staab et al. (2017) to $1\,\sigma$. Similar activity levels are observed in the Pleiades cluster, for stars aged under 130 Myr (Mamajek & Hillenbrand, 2008). Equation 9 from Mascareño et al. (2015) (which is based on the empirical correlation between $P_{\rm rot}$ and $\log(R'_{\rm HK})$ in a HARPS population of non-HJ hosts) predicts WASP-43 to have $\log(R'_{\rm HK}) = -4.62 \pm 0.07$.

To rule out that high activity is not a symptom of young age, Esposito et al. (2017) calculated space velocities using spectroscopic parallaxes reported in Hellier et al. (2011). They found WASP-43's space velocity not to be compatible with any known nearby young moving groups (reported in Zuckerman & Song, 2004), concluding that enhanced activity likely results from tidal or magnetic SPI. However, they find no evidence of stellar rotational spin-up. The anomalously high $\log(R'_{\rm HK})$ measurements have been caused by stellar flares: dramatic short term variability has been seen in a young, rapidly rotating planet hosts (Hernán-Obispo et al., 2010, 2015). Staab et al. (2017) find a $< 14\%$ probability of observing a flare during exposure in systems analogous to WASP-12. Magnetic SPI could stimulate a higher occurrence rate of flares.

Second observation

I observed WASP-43 at phase $\phi = 0.671$, finding a similarly high activity value of $\log(R'_{\rm HK}) = -4.18^{+0.09}_{-0.11}$. This provides a $3\,\sigma$ level of confidence of enhanced chromospheric activity compared to the $\log(R'_{\rm HK}) = -4.62 \pm 0.07$ prediction by Equation 9 in Mascareño et al. (2015). It is unlikely that two observations made at a considerable time apart would observe flaring events that affect activity almost identically. Moreover, Esposito et al. (2017) report an average $\log(R'_{\rm HK})$ value, so presumably made multiple observations.

WASP-43 is an extreme system for both tidal and magnetic SPI proxies, as shown in Figure 4.13. It has semi-major axis $a = 0.015$ AU, while the orbital angular momentum of the planet is higher than the rotational angular momentum of the host (see Table 5.2). If the planet has migrated, or is still migrating inward, a sufficient budget of angular momentum may spin-up the host or inhibit stellar spin-down from occurring at a typical rate (Poppenhaeger, 2015). SPMI proxy M_P/a^2 is $\sim$ 20 times the value defined by Miller et al. (2015) as representing extreme SPI systems. The anomalously high activity of WASP-43 is likely caused by extreme interactions with its close-orbiting companion.

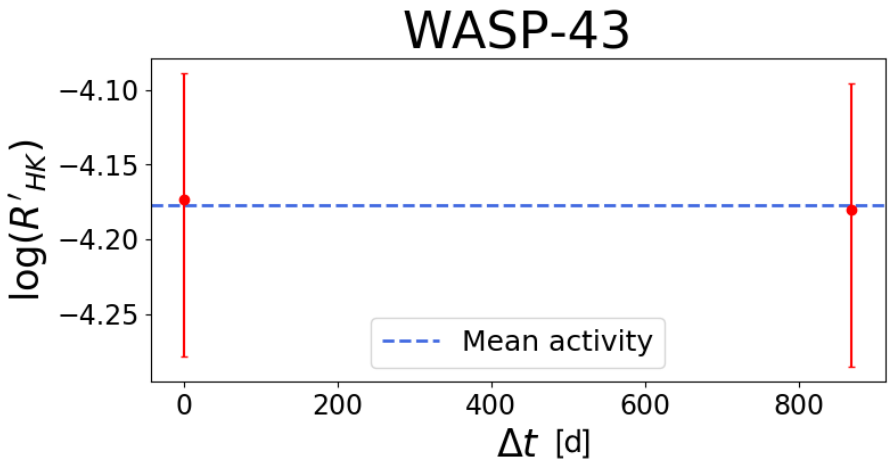

FIGURE 4.14: Observations of WASP-43 taken more than 800 days apart.

4.6.2 WASP-103

Extreme HJ WASP-103b is in an ultra-short (0.93 d) period orbit around a V = 12.1, F8V star (Gillon et al., 2014). It has semi-major axis a = 0.02 AU, which is only $\sim$ 1.16 times its Roche limit—making it close to tidal disruption. It is approximately 1.5 M_J, with an inflated 1.65 R_J radius that fills 54% of its Roche lobe. Incident flux is $>$ 6000 times that received by Earth, making WASP-103b one of the most highly irradiated known HJs. It has an equilibrium temperature $\sim$ 2500 K and surface gravity $\sim$ 14 ms^{-2}. After WASP-18, it has the second largest SPTI proxy value in the OU-SALT sample ($h_{\rm scale}/h_{\rm tide} = 0.93$).

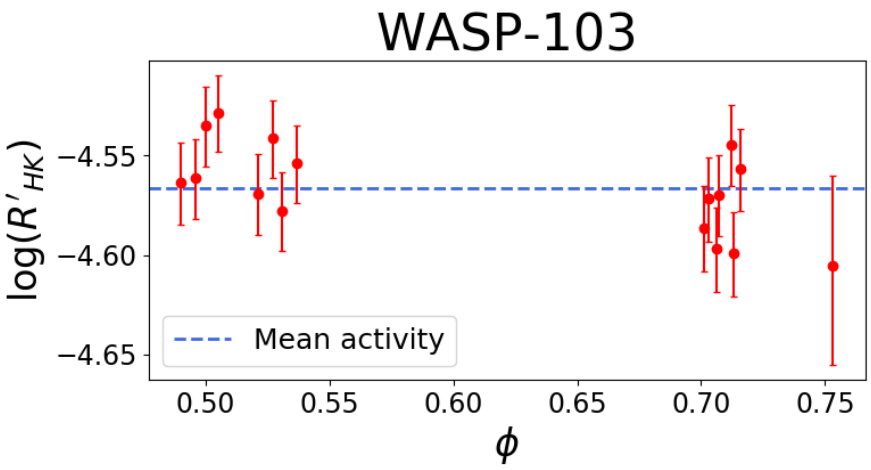

FIGURE 4.15: Activity-phase plot for WASP-103 with a 12-day baseline for all observations.

I found the median activity of WASP-103 to be $\log(R'_{\rm HK}) = -4.57 \pm 0.04$, making it an active star. Figure 4.15 indicates that WASP-103's activity is relatively stable, though observations were made within limited phase windows. WASP-103 might be expected to display WASP-12-like depressed emission due to the extreme nature of the planetary system and large Roche-filling factor (Staab et al., 2017). SPTI may have spun-up the host star, such that extraordinarily high activity is masked by mass-loss. However, WASP-103b's estimated orbital angular momentum ($J_{\rm orb} = 1.66 \times 10^{47}$) is less than the rotational angular momentum of the host ($J_\star = 8.67 \times 10^{47}$), making the spin-up scenario less likely. An alternative explanation is required.

A stable, azimuthally symmetric disk of ablated material is required to depress observed activity (Haswell et al., 2012). As noted in Section 4.4.2, radiation incident upon WASP-103b is extreme. Solar wind and radiative pressure may thus be sufficient around the massive, hot host that ablated material is destabilised, either accreting onto the star or being blown from the system. A further destabilising factor may involve intense interactions between the magnetic fields of the planet and star. Thus, although mass loss may be occurring, the ecosystem required for the formation of a stable gas disk is not *in situ*. Alternatively, the strong stellar wind may inhibit atmospheric escape (Vidotto & Cleary, 2020).

4.6.3 KELT-11

KELT-11b is in a 4.7 d day orbit around bright ($V = 8$), evolved subgiant HD 93396, with $T_{\rm eff} = 5370$ K and $M_\star = 1.44\,M_\odot$ (Pepper et al., 2017). It is an inflated sub-Saturn planet with $M_{\rm P} = 0.2\,M_{\rm J}$, $R_{\rm P} = 1.37\,R_{\rm J}$, $\rho = 0.093$ g cm^{-3}, $T_{\rm eq} = 1712$ K, and an exceptionally large atmospheric scale height of 2763 km. I have found the host to have relatively depressed median activity ($\log(R'_{\rm HK}) = -5.018^{+0.046}_{-0.052}$) that may result from its evolutionary phase and/or from mass loss. Figure 4.16 shows activity varying with an amplitude of $\sim$ 0.09 over $\sim$ 100 d. Activity variation is more likely to occur in highly-active stars, thus this level of variation is somewhat unexpected. It may result from natural activity cycles, or through interactions with a companion. SPI with KELT-11b is unlikely to cause the variation given the apparent $\sim$ 100 d period. Figure 4.16 confirms that the variation is not modulated to KELT-11b's orbital period.

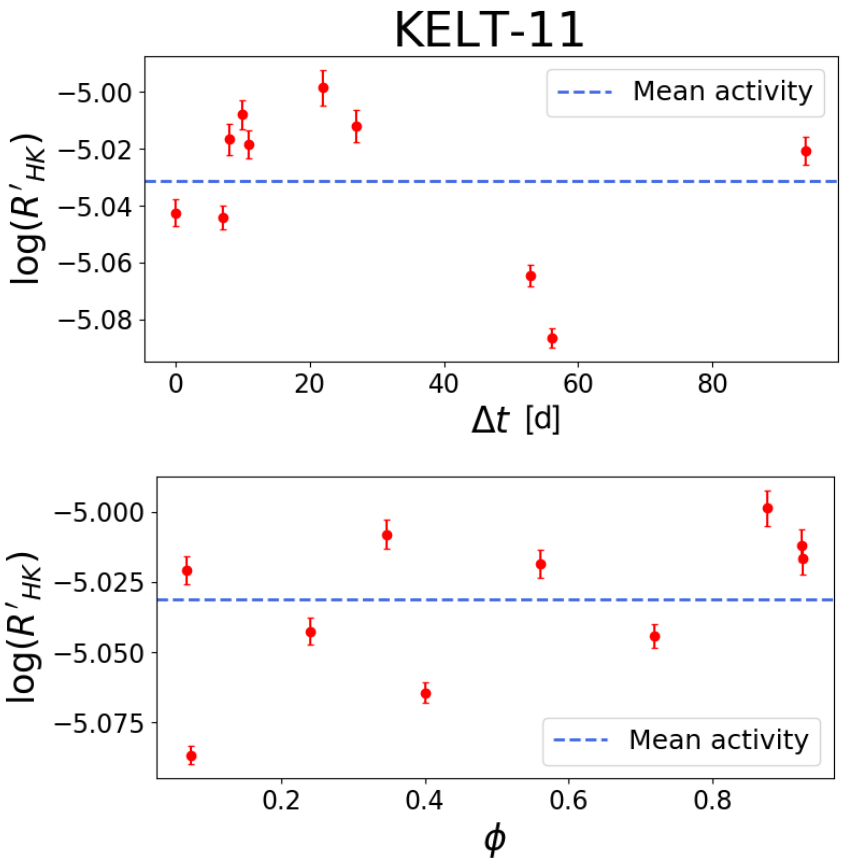

FIGURE 4.16: Top panel: time in days since initial observation plotted against activity for KELT-11. Lower panel: the variation is not obviously phase-modulated.

4.6.4 HD 73256

HD 73256b is a 1.85 M_J HJ in a 2.6 d orbital period around a bright ($V = 8.08$) G8/K0 dwarf (Udry et al., 2003). The planet does not transit, so ablated material does not necessary lie in the observation plane. It is relatively cool ($T_{eq} = 1318$ K). HD 73256's median activity is observed to be $\log(R'_{HK}) = -4.417^{+0.042}_{-0.028}$, making it an active MS star. Variations in activity of amplitude ~ 0.04 dex are present in Figure 4.17. Further observations are required to confirm whether these variations result from SPTI with the HJ companion.

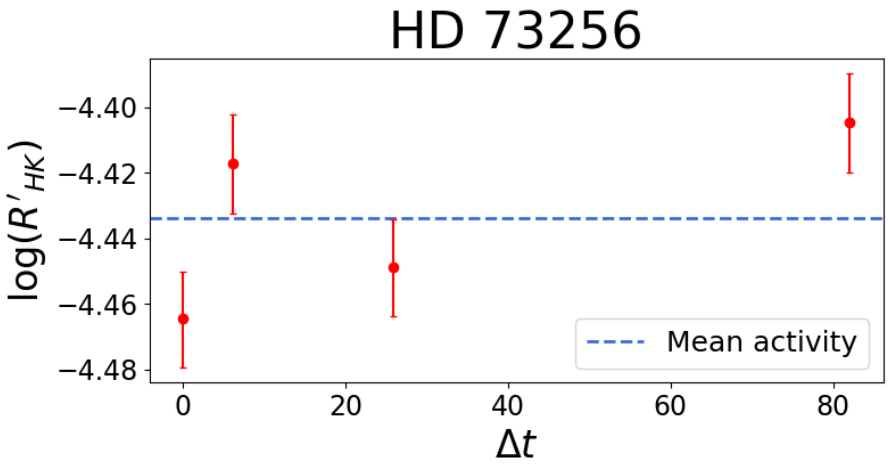

FIGURE 4.17: Time in days since initial observation plotted against activity for HD 73256.

4.6.5 Other systems

Figure 4.18 shows time-series plots for the remaining OU-SALT targets with multiple observations. Of note, K2-32 (Sinukoff et al., 2016), WASP-63 (Hellier et al., 2012), WASP-67 (Hellier et al., 2012) and WASP-104 (Smith, A. M. S. et al., 2014) show activity variation with amplitude $> 1\,\sigma$ uncertainties, making them priority targets for follow-up observations.

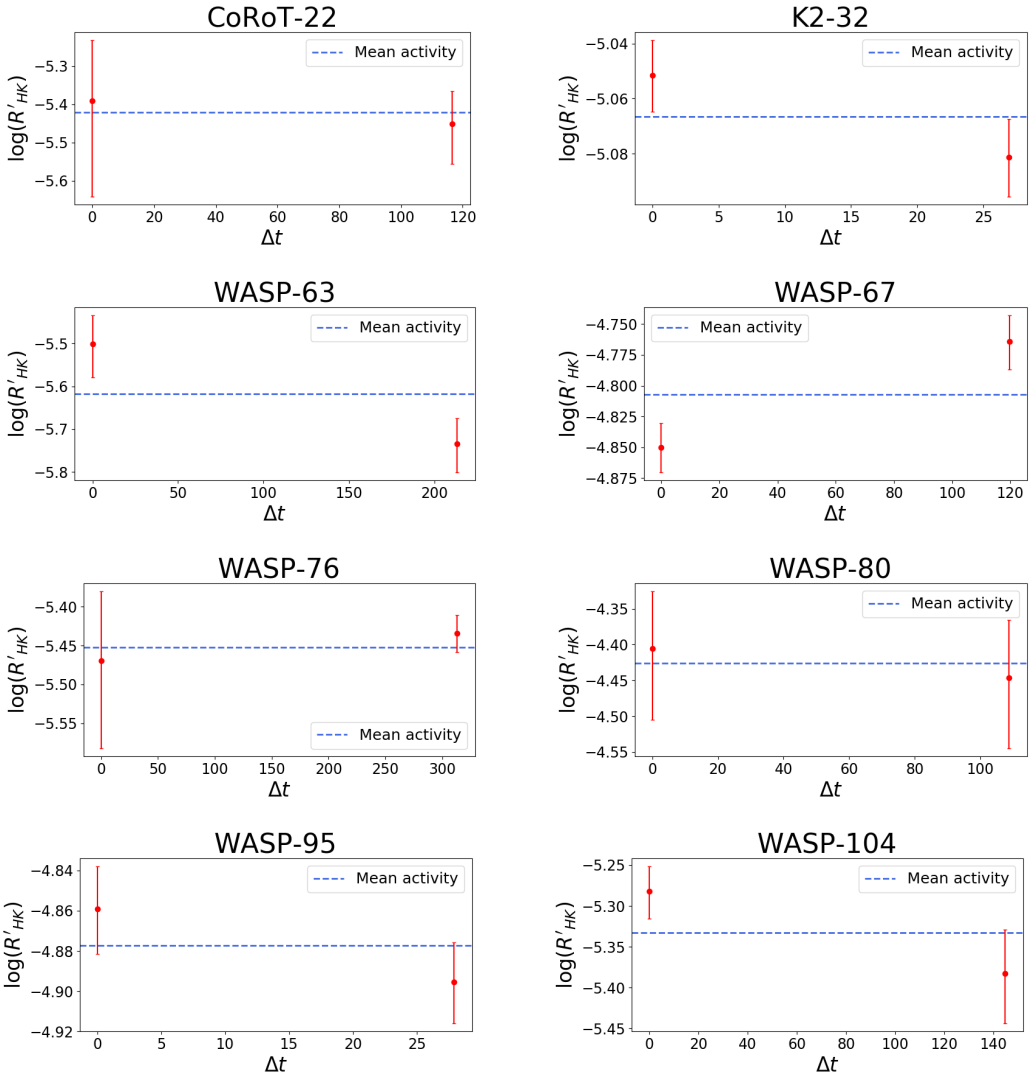

FIGURE 4.18: Activity variation over time elapsed since first observation for targets where multiple observations have been made. Dashed blue line shows mean host activity.

4.7 A novel search for orbitally-modulated SPI

Our observations are snap-shots of activity at specific planetary orbital phases. Figure 4.19 shows orbital phase (ϕ) plotted against activity for all observations. The sample is divided into high and low activity sub-samples either side of $\log(R'_{\rm HK}) = -4.9$ (see Section 5.2 for justification of this division). The OU-SALT sample is therefore treated as one high activity and one low activity target. A seven-point moving average is plotted to track phase-modulated activity variation. A dozen binned points are plotted for each sub-sample and fitted with a sinusoidal curve.

Periodicity is evident in each sub-sample with amplitude ~ 0.25 and peaks at $\phi \sim 0.2$ and 0.8 (i.e. approximately at quadrature) in the high activity population, and at $\phi \sim 0.4$ and 0.8 in the low activity population (where activity signal may be masked by enshrouding). Shkolnik et al. (2003, 2005, 2008) similarly identify chromospheric activity peaks for HD 179949 at a planetary orbital phase of $\phi \sim 0.75$. As there was not a second peak offset by 180°, these observations were interpreted as magnetic SPI. Following reconnection events, particles are accelerated along field lines, creating hotspots where they slam into the chromosphere (Cuntz et al., 2000; Lanza, 2008). Star-planet magnetic connectivity is complex. The presence of the planet's magnetosphere near the star prevents expansion of the coronal magnetic field and the acceleration of the stellar wind. The pressure gradient is smaller, so the coronal field lines that would be opened by the wind remain closed and the plasma in these loops does not escape. This may result in observable phase shifts of the hot spots from the star–planet line (Cohen et al., 2009). Saar et al. (2004) suggest that HD 179949's emission excess occurs $\Delta\phi = 0.17$ ahead of the sub-planet point of the star.

Figure 4.19 indicates SPI in action in the OU-SALT sample as a whole, with hot spots occurring at $\phi \approx 0.8$. However, a chi-squared value of $\chi^2 = 0.039$ for the fit of the sinusoid to the high activity binned points does not provide strong evidence for rejecting the null hypothesis. More observations are required.

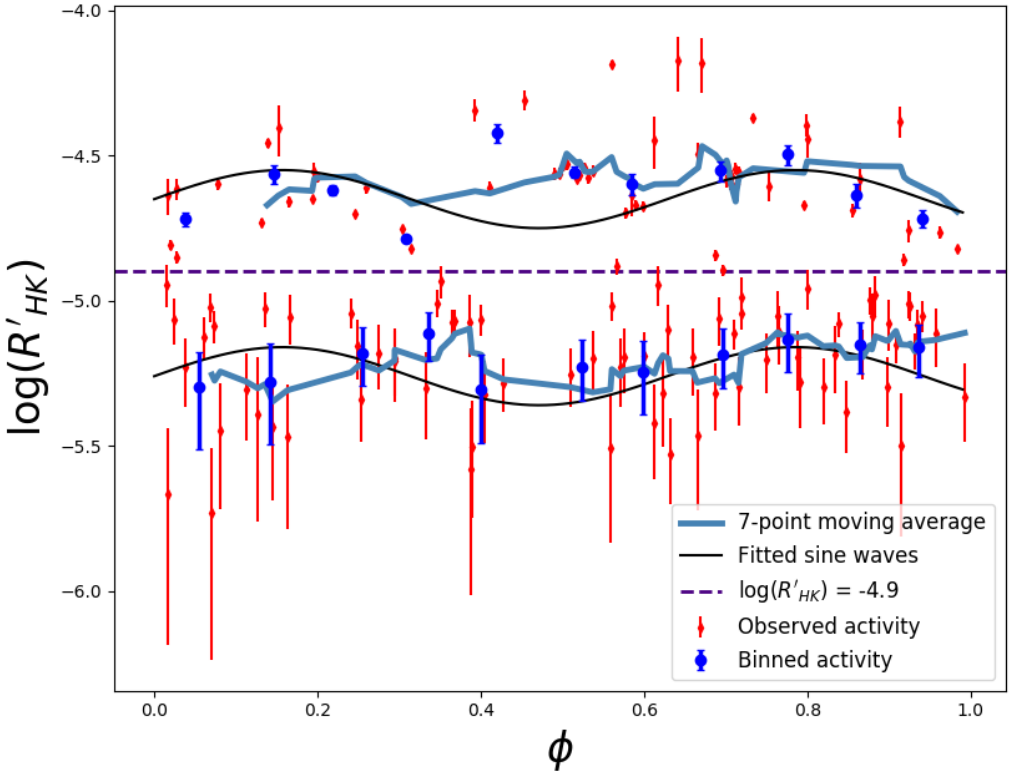

FIGURE 4.19: Activity phase plot for 141 observations. The OU-sample has been divided into high ($\log(R'_{\rm HK}) < -4.9$) and low ($\log(R'_{\rm HK}) > -4.9$) activity sub-samples. Red points show observed activity (with instrumental errors). Blue points show observations in 12 bins. Sinusoids are fitted for the low and high activity sub-samples.

TABLE 4.3: Planetary parameters. From TEPCAT: orbital period $P_{\rm orb}$, semi-major axis a, planet mass $M_{\rm P}$, planet radius $R_{\rm P}$. Calculated: incident flux $S_{\rm P}$ in Earth units and percentage Roche-lobe filling factor $V_{\rm Roche}$.

System	$P_{\rm orb}$ [days]	a [AU]	$M_{\rm P}$ [$M_{\rm J}$]	$R_{\rm P}$ [$R_{\rm J}$]	$S_{\rm P}$ [$U_{\oplus}$]	$V_{\rm Roche}$ (%)
CoRoT-7	0.85	0.017	$0.02^{+0.003}_{-0.003}$	$0.14^{+0.006}_{-0.006}$	1590±80	2.8
CoRoT-11	2.99	0.044	$2.34^{+0.39}_{-0.39}$	$1.43^{+0.06}_{-0.06}$	1420±163	1.9
CoRoT-22	9.76	0.092	$0.06^{+0.09}_{-0.02}$	$0.43^{+0.02}_{-0.04}$	170±27	0.2
CoRoT-28	5.21	0.059	$0.48^{+0.09}_{-0.09}$	$0.95^{+0.07}_{-0.07}$	575±88	0.9
EPIC 219388192	5.29	0.059	$36.84^{+0.97}_{-0.97}$	$0.94^{+0.03}_{-0.03}$	305±35	0.01
HAT-P-27	3.04	0.040	$0.66^{+0.03}_{-0.03}$	$1.04^{+0.08}_{-0.06}$	356±44	2.5
HATS-2	1.35	0.023	$1.34^{+0.15}_{-0.15}$	$1.17^{+0.03}_{-0.03}$	1020±51	9.0
HATS-3	3.55	0.049	$1.07^{+0.14}_{-0.14}$	$1.38^{+0.04}_{-0.04}$	1220±60	2.7
HATS-10	3.31	0.045	$0.53^{+0.08}_{-0.08}$	$0.97^{+0.06}_{-0.05}$	650±68	2.1
HATS-13	3.04	0.041	$0.54^{+0.07}_{-0.07}$	$1.21^{+0.04}_{-0.04}$	399±19	4.8
HATS-18	0.84	0.018	$1.98^{+0.08}_{-0.08}$	$1.34^{+0.10}_{-0.05}$	2960±343	25
HATS-21	3.55	0.047	$0.33^{+0.04}_{-0.03}$	$1.12^{+0.15}_{-0.05}$	450±79	4.5
HATS-27	4.64	0.061	$0.53^{+0.13}_{-0.13}$	$1.50^{+0.20}_{-0.11}$	1250±246	4.0
HATS-29	4.61	0.055	$0.65^{+0.06}_{-0.06}$	$1.25^{+0.06}_{-0.06}$	356±28	1.9
HATS-30	3.17	0.044	$0.71^{+0.04}_{-0.04}$	$1.18^{+0.05}_{-0.05}$	665±51	3.1
HATS-36	4.17	0.054	$3.22^{+0.06}_{-0.062}$	$1.24^{+0.04}_{-0.04}$	582±40	0.5
HD 73256	2.55	0.037	$1.87^{+0.49}_{-0.49}$		500	
K2-2	9.12	0.098	$0.04^{+0.004}_{-0.004}$	$0.23^{+0.02}_{-0.02}$	32±3	0.03
K2-19	7.92	0.076	$0.09^{+0.02}_{-0.02}$	$0.67^{+0.07}_{-0.07}$	99±11	0.7
K2-31	1.26	0.022	$1.86^{+0.08}_{-0.08}$	$1.06^{+0.35}_{-0.35}$	1550±335	6.3
K2-32	8.99	0.080	$0.05^{+0.01}_{-0.01}$	$0.46^{+0.03}_{-0.03}$	77±8	0.3
K2-39	4.61	0.057	$0.12^{+0.01}_{-0.01}$	$0.51^{+0.06}_{-0.06}$	1360±203	0.7
K2-96	0.96	0.018	$0.02^{+0.001}_{-0.001}$	$0.15^{+0.02}_{-0.01}$	1540±149	3.1
K2-106	0.57	0.013	$0.02^{+0.003}_{-0.003}$	$0.14^{+0.01}_{-0.01}$	4110±441	3.9
K2-135	1.21	0.021	$0.01^{+0.002}_{-0.002}$	$0.14^{+0.02}_{-0.01}$	262±83	1.8
KELT-10	4.17	0.052	$0.68^{+0.04}_{-0.04}$	$1.4^{+0.07}_{-0.05}$	596±51	3.2
KELT-11	4.74	0.062	$0.17^{+0.02}_{-0.02}$	$1.35^{+0.10}_{-0.10}$	1400±57	8.6
KELT-15	3.33	0.046	$1.20^{+0.07}_{-0.07}$	$1.52^{+0.12}_{-0.11}$	1200±152	3.6
Qatar-2	1.34	0.021	$2.47^{+0.06}_{-0.06}$	$1.11^{+0.01}_{-0.01}$	453±14	4.5
WASP-5	1.63	0.027	$1.59^{+0.05}_{-0.05}$	$1.18^{+0.06}_{-0.06}$	1570±126	5.3
WASP-6	3.36	0.041	$0.48^{+0.03}_{-0.03}$	$1.23^{+0.04}_{-0.04}$	326±26	4.6
WASP-7	4.96	0.062	$0.98^{+0.13}_{-0.13}$	$1.37^{+0.09}_{-0.09}$	910±113	1.5
WASP-8	8.16	0.080	$2.25^{+0.08}_{-0.08}$	$1.04^{+0.01}_{-0.05}$	123±14	0.1
WASP-15	3.75	0.052	$0.59^{+0.02}_{-0.02}$	$1.41^{+0.05}_{-0.05}$	1460±92	4.5

System	P_{orb}	a	M_P	R_P	S_P	V_{Roche}
	[days]	[AU]	[M_J]	[R_J]	[$U_\oplus$]	(%)
WASP-17	3.73	0.051	$0.48^{+0.03}_{-0.03}$	$1.93^{+0.05}_{-0.05}$	1570±103	14.8
WASP-18	0.94	0.021	$10.52^{+0.32}_{-0.32}$	$1.2^{+0.03}_{-0.03}$	5620±303	2.7
WASP-25	3.77	0.048	$0.6^{+0.05}_{-0.05}$	$1.25^{+0.03}_{-0.03}$	357±16	3.1
WASP-26	2.77	0.040	$1.02^{+0.03}_{-0.03}$	$1.22^{+0.05}_{-0.05}$	1230±77	3.2
WASP-28	3.41	0.044	$0.89^{+0.06}_{-0.06}$	$1.22^{+0.03}_{-0.03}$	739±48	2.4
WASP-32	2.72	0.039	$3.6^{+0.07}_{-0.07}$	$1.18^{+0.07}_{-0.07}$	987±90	0.9
WASP-34	4.32	0.052	$0.59^{+0.01}_{-0.01}$	$1.22^{+0.11}_{-0.08}$	299±77	2.2
WASP-36	1.54	0.027	$2.36^{+0.07}_{-0.07}$	$1.33^{+0.02}_{-0.02}$	1500±53	5.9
WASP-38	6.87	0.075	$2.69^{+0.06}_{-0.06}$	$1.09^{+0.03}_{-0.03}$	402±20	0.1
WASP-43	0.81	0.015	$2.03^{+0.05}_{-0.05}$	$1.04^{+0.02}_{-0.02}$	716±29	11.8
WASP-44	2.42	0.034	$0.87^{+0.08}_{-0.08}$	$1.00^{+0.04}_{-0.04}$	481±53	2.7
WASP-45	3.13	0.041	$1.00^{+0.06}_{-0.06}$	$0.99^{+0.04}_{-0.04}$	311±22	1.4
WASP-46	1.43	0.023	$1.91^{+0.13}_{-0.13}$	$1.17^{+0.04}_{-0.04}$	1190±105	5.8
WASP-47	4.16	0.051	$1.14^{+0.02}_{-0.02}$	$1.12^{+0.01}_{-0.01}$	430±13	1.0
WASP-50	1.96	0.029	$1.44^{+0.07}_{-0.07}$	$1.14^{+0.03}_{-0.03}$	717±45	3.7
WASP-51	2.81	0.042	$0.71^{+0.03}_{-0.03}$	$1.34^{+0.07}_{-0.07}$	1220±106	5.8
WASP-52	1.75	0.026	$0.43^{+0.02}_{-0.02}$	$1.25^{+0.03}_{-0.03}$	496±29	20.7
WASP-55	4.47	0.056	$0.63^{+0.04}_{-0.04}$	$1.33^{+0.03}_{-0.02}$	475±21	2.6
WASP-62	4.41	0.057	$0.57^{+0.04}_{-0.04}$	$1.39^{+0.06}_{-0.06}$	689±56	3.3
WASP-63	4.38	0.057	$0.38^{+0.03}_{-0.03}$	$1.43^{+0.10}_{-0.06}$	1030±112	5.4
WASP-65	2.31	0.033	$1.55^{+0.16}_{-0.16}$	$1.11^{+0.06}_{-0.06}$	807±115	2.3
WASP-67	4.61	0.051	$0.41^{+0.04}_{-0.04}$	$1.09^{+0.05}_{-0.05}$	198±15	2.0
WASP-69	3.87	0.045	$0.26^{+0.02}_{-0.02}$	$1.06^{+0.05}_{-0.05}$	141±10	4.1
WASP-70	3.71	0.049	$0.59^{+0.02}_{-0.02}$	$1.16^{+0.07}_{-0.10}$	594±88	2.6
WASP-72	2.22	0.037	$1.46^{+0.06}_{-0.06}$	$1.27^{+0.20}_{-0.20}$	3910±953	3.9
WASP-73	4.09	0.055	$1.88^{+0.07}_{-0.06}$	$1.16^{+0.12}_{-0.08}$	1670±309	0.7
WASP-74	2.14	0.037	$0.95^{+0.06}_{-0.06}$	$1.56^{+0.06}_{-0.06}$	2270±185	12.1
WASP-75	2.48	0.038	$1.08^{+0.05}_{-0.05}$	$1.31^{+0.02}_{-0.02}$	1410±63	4.6
WASP-76	1.81	0.033	$0.92^{+0.03}_{-0.03}$	$1.83^{+0.06}_{-0.04}$	3770±208	29.0
WASP-77	1.36	0.024	$1.76^{+0.06}_{-0.06}$	$1.21^{+0.02}_{-0.02}$	1400±61	7.7
WASP-78	2.17	0.037	$0.86^{+0.08}_{-0.08}$	$2.06^{+0.10}_{-0.10}$	5100±487	30.8
WASP-79	3.66	0.052	$0.86^{+0.08}_{-0.08}$	$1.53^{+0.04}_{-0.04}$	1440±88	4.2
WASP-80	3.07	0.035	$0.56^{+0.03}_{-0.03}$	$0.99^{+0.02}_{-0.02}$	77±4	2.5
WASP-82	2.71	0.045	$1.25^{+0.05}_{-0.05}$	$1.71^{+0.09}_{-0.06}$	3940±367	7.5
WASP-87 A	1.68	0.029	$2.18^{+0.15}_{-0.15}$	$1.39^{+0.06}_{-0.06}$	4740±434	6.0
WASP-88	4.95	0.064	$0.56^{+0.08}_{-0.08}$	$1.7^{+0.13}_{-0.07}$	1610±188	4.8
WASP-89	3.36	0.043	$5.9^{+0.40}_{-0.40}$	$1.04^{+0.04}_{-0.04}$	230±20	0.2

System	P_{orb}	a	M_P	R_P	S_P	V_{Roche}
	[days]	[AU]	[M_J]	[R_J]	[$U_\oplus$]	(%)
WASP-90	3.92	0.056	$0.63^{+0.07}_{-0.07}$	$1.63^{+0.09}_{-0.09}$	1920±193	6.1
WASP-91	2.80	0.037	$1.34^{+0.08}_{-0.08}$	$1.03^{+0.04}_{-0.04}$	284±25	1.4
WASP-94 A	3.95	0.055	$0.45^{+0.04}_{-0.03}$	$1.72^{+0.06}_{-0.05}$	1130±81	10.0
WASP-94 B	2.01	0.034	$0.62^{+0.03}_{-0.03}$		1940±352	
WASP-95	2.19	0.034	$1.13^{+0.10}_{-0.04}$	$1.21^{+0.06}_{-0.06}$	1140±170	4.5
WASP-96	3.43	0.045	$0.48^{+0.03}_{-0.03}$	$1.2^{+0.06}_{-0.06}$	441±49	4.2
WASP-97	2.07	0.033	$1.32^{+0.05}_{-0.05}$	$1.13^{+0.06}_{-0.06}$	956±79	3.5
WASP-99	5.75	0.072	$2.78^{+0.13}_{-0.13}$	$1.1^{+0.08}_{-0.05}$	774±103	0.2
WASP-100	2.85	0.046	$2.03^{+0.12}_{-0.12}$	$1.69^{+0.29}_{-0.29}$	3900±1182	4.0
WASP-101	3.59	0.051	$0.5^{+0.04}_{-0.04}$	$1.41^{+0.05}_{-0.05}$	967±69	5.9
WASP-103	0.93	0.020	$1.47^{+0.11}_{-0.13}$	$1.65^{+0.05}_{-0.06}$	6380±603	54.1
WASP-104	1.75	0.029	$1.31^{+0.05}_{-0.05}$	$1.11^{+0.02}_{-0.02}$	856±40	4.6
WASP-105	7.87	0.075	$1.8^{+0.10}_{-0.10}$	$0.96^{+0.03}_{-0.03}$	85±9	0.1
WASP-108	2.68	0.040	$0.89^{+0.06}_{-0.06}$	$1.28^{+0.05}_{-0.05}$	1090±90	4.5
WASP-109	3.32	0.046	$0.91^{+0.13}_{-0.13}$	$1.44^{+0.05}_{-0.05}$	1370±111	4.1
WASP-111	2.31	0.039	$1.83^{+0.15}_{-0.15}$	$1.44^{+0.09}_{-0.09}$	3370±401	4.2
WASP-114	1.55	0.029	$1.77^{+0.06}_{-0.06}$	$1.34^{+0.06}_{-0.06}$	2810±248	7.9
WASP-117	10.02	0.095	$0.28^{+0.01}_{-0.01}$	$1.02^{+0.08}_{-0.07}$	183±21	0.5
WASP-118	4.05	0.054	$0.52^{+0.18}_{-0.18}$	$1.39^{+0.01}_{-0.01}$	1570±40	4.3
WASP-119	2.50	0.036	$1.23^{+0.08}_{-0.08}$	$1.4^{+0.20}_{-0.20}$	1000±171	4.9
WASP-120	3.61	0.051	$4.85^{+0.21}_{-0.21}$	$1.47^{+0.10}_{-0.10}$	2060±248	0.7
WASP-122	1.71	0.030	$1.28^{+0.03}_{-0.03}$	$1.74^{+0.05}_{-0.05}$	2460±110	20.0
WASP-123	2.98	0.043	$0.9^{+0.04}_{-0.04}$	$1.32^{+0.07}_{-0.07}$	886±77	3.9
WASP-126	3.29	0.045	$0.28^{+0.04}_{-0.04}$	$0.96^{+0.10}_{-0.05}$	813±131	3.9
WASP-127	4.18	0.052	$0.18^{+0.02}_{-0.02}$	$1.37^{+0.03}_{-0.03}$	701±33	11.1
WASP-130	11.55	0.101	$1.23^{+0.04}_{-0.04}$	$0.89^{+0.03}_{-0.03}$	79±5	0.1
WASP-131	5.32	0.061	$0.27^{+0.02}_{-0.02}$	$1.22^{+0.05}_{-0.05}$	715±51	3.2
WASP-132	7.13	0.067	$0.41^{+0.03}_{-0.03}$	$0.87^{+0.03}_{-0.03}$	56±3	0.4
WASP-138	3.63	0.049	$1.22^{+0.08}_{-0.08}$	$1.09^{+0.05}_{-0.05}$	1070±84	1.1
WASP-139	5.92	0.062	$0.12^{+0.02}_{-0.018}$	$0.8^{+0.05}_{-0.05}$	118±14	1.7
WASP-140	2.24	0.032	$2.44^{+0.07}_{-0.07}$	$1.44^{+0.42}_{-0.18}$	514±50	3.4
WASP-142	2.05	0.035	$0.84^{+0.09}_{-0.09}$	$1.53^{+0.08}_{-0.08}$	2710±286	14.1
WASP-156	3.84	0.045	$0.13^{+0.01}_{-0.01}$	$0.51^{+0.02}_{-0.02}$	147±13	0.9

4.8 Summary

This chapter has explored the OU-SALT sample. Target selection criteria were discussed and justified. Defining the evolutionary status of OU-SALT targets is of key importance as main sequence and evolved stars have different activity distributions. Absolute magnitude was calculated using Gaia DR2 distances deduced with Bayesian techniques by Bailer-Jones et al. (2018). The OU-SALT targets are at distances between 30–900 pc.

HR diagrams plotted with metallicity categories and an empirical main sequence highlighted the importance of taking metallicity into account when categorising targets. The approach taken by previous studies to distinguish between main sequence and evolved stars has been improved upon by exploiting the precision of Gaia DR2 data. The Terminal Age Main Sequence (TAMS), identified from the Padova catalogue, was plotted along with target Gaia DR2 effective temperatures and absolute magnitudes to distinguish between main sequence (MS) and (SG) targets. The OU-SALT population has thus been divided into 84 MS targets and 20 SGs, with a high level of confidence in the MS population.

The ages of the systems have been considered, with activity and gyro ages both proving unreliable for these systems. Dating the systems with isochrones may be preferable.

The demographics of the OU-SALT population were explored. The targets tend to stack up on the upper boundary of the Neptune Desert (ND: gas giants) or be located in and around the lower boundary (super Earths). Several key aspects of the planetary environment have been considered. The flux incident on the targets ranges between $30 - 6400\ \mathrm{U}_{\oplus}$, making these highly irradiated systems. Roche-lobe filling factors are calculated and, perhaps surprisingly, no planets are so distorted as to overfill their Roche lobes. WASP-103 presents the highest filling factor of 54%.

Several SPI proxies were calculated. The half of the sample for which $J_{\mathrm{orb}} \gtrsim J_{\star}$ are priority candidates for SPTI studies. Approximately half of the sample also qualify as "extreme' systems" ($M_{\mathrm{P}}/a^2 > 450\ M_{\mathrm{J}}\ \mathrm{AU}^{-2}$: Miller et al., 2015). The OU-SALT sample is an ideal test-bed for the search for SPI.

Times-series activity variations for 13 systems were presented. WASP-43 was confirmed as a planet-host with anomalously high activity ($\log(R'_{\rm HK}) = -4.18^{+0.09}_{-0.11}$). WASP-103 was also identified as a high-activity host ($\log(R'_{\rm HK}) = -4.57 \pm 0.04$). It does not provide the ideal conditions for the formation of an absorbing, diffuse circumstellar gas cloud. A $\sim$ 100 d activity (sub-)cycle was potentially identified for KELT-11. Significant activity variations were identified in five other systems.

Finally, I took a novel approach to search for systematic activity variations in the sample as a whole, effectively treating the OU-SALT targets as two hosts. Sinusoidal-like variations of amplitude 0.25 dex, with peaks at $\phi \sim 0.2$ and 0.8, were identified in the high activity sample. Emission offsets from the sub-planet point on the star are predicted (Cohen et al., 2009). These results possibly indicate orbitally-modulated SPI in action but are not conclusive.

Having classified and considered the physical environments of these systems, we are able to study the correlations between stellar/planetary variables and activity in context.

Collaboration details

Section 4.4.3 draws on the work of Busuttil (2017). I set-up and tested the environment in which the Roche equipotential code detailed was run, and provided inputs for the OU-SALT targets. Richard Busuttil wrote and ran the code, and generated the plots in Section 4.4.3.

Chapter 5

OU-SALT population study

In Section 5.1, I comment on the scope of this population study. I then place OU-SALT targets into the context of published activity catalogues in Section 5.2, before carefully considering the potential influence of absorption in the interstellar medium on our results (Section 5.3). A systematic study of the relationship between planetary variables and activity is undertaken in Section 5.4, which is extended to SPI terms and proxies in Section 5.5. I address the relationship between stellar parameters and activity in Section 5.6, which provides vital context for understanding planet-activity correlations. I address interdependencies between variables in Section 5.7. I then attempt to constrain systems in which significant absorption of emission cores occurs in Section 5.8. I provide a summary and details of collaboration in Section 5.9.

Findings in this chapter were presented at the Twentieth Cambridge Workshops of Cool Stars, Stellar Systems and the Sun, Boston 2019[1].

[1] http://coolstars20.cfa.harvard.edu/abstracts.html

5.1 Scope of study

In this chapter, I study the OU-SALT sample as a whole to consider whether the presence of close-in planets has a ubiquitous influence on stellar activity. Single epoch observations facilitate the study of comparatively long-term effects rather than those that vary on shorter time-frames, such as a planet's orbital period. In particular, I focus on the relationship between activity and planetary parameters. The OU-SALT dataset is unique in activity work, both in its homogeneity and due to the extreme nature of the sample objects (mainly hot Jupiter hosts). This activity study therefore provides unprecedented insights into mass-loss, enshrouding, absorption, star-planet interactions (SPI), evaporation deserts, and planetary evolution.

5.2 Activity in context

To place the OU-SALT sample in context, I compare it to two activity catalogues:

Boro-Saikia et al. (2018)

Boro Saikia et al. (2018) (BS18) have compiled a comprehensive catalogue of the chromospheric activity of 4454 cool stars from a combination of archival High Accuracy Radial velocity Planet Searcher (HARPS: Lovis et al., 2011b) spectra and multiple other surveys, including the Mount Wilson survey (Figure 5.1). The sample thus contains activity measurements for both planet hosts and stars around which planets have not been discovered. Figure 5.2(a) shows a plot of colour and activity values published in BS18, with the position of the suggested Vaughan-Preston (VP) gap indicated (Foukal, 2018). The peak of the activity distribution is within the VP gap. This leads BS18 to argue that the VP gap is less prominent than previously thought.

I have identified several issues with the conclusions reached in this paper. Figure 1 of BS18 shows that the sample is composed of MS stars. With the benefit of Gaia DR2 data, I re-plot the HR diagram for BS18 in Figure 5.1 along with the terminal age main sequence (TAMS) for stars with solar metallicity. The BS18 sample is contaminated with a number of evolved stars.

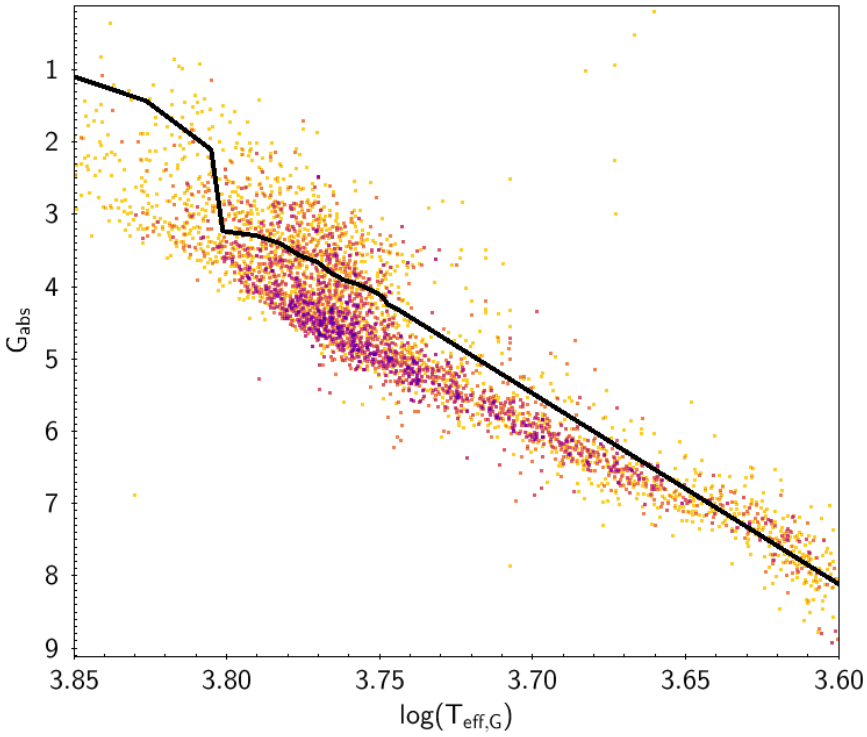

FIGURE 5.1: HR diagram of the BS18 sample, plotted with Gaia DR2 measurements.

The Mount Wilson system is only calibrated in the range $0.4 < B-V < 1.2$ (Noyes et al., 1984; Staab et al., 2017). However, BS18 measure $\log(R'_{HK})$ for stars with $0 < B-V < 2.0$. The applicability of $\log(R'_{HK})$'s is extended to K and M stars using the polynomial scaling relationship detailed in Equation 3 of BS18, referring to a seemingly unpublished paper (Marvin et al. 2018 in prep). BS18 argue that the basal limit increases for redder stars. However, the increase in $\log(R'_{HK})$ values may also be attributed to lower flux in the red passband.

Most strikingly, there is a distinct boundary feature (indicated with a dashed red line) between high and low density regions in in Figure 5.2(a). One expects to observe such a feature at the basal limit. This suggests the S to $\log(R'_{HK})$ calibration has been undertaken incorrectly. I recalculated $\log(R'_{HK})$ from S values in BS18 (using the calibration relationship detailed in Equation 3.2), and plot these results in Figure 5.2(b). The feature is still present but now parallel to the x-axis and located at $\log(R'_{HK}) \sim -5.1$. The activity peak has shifted beneath the VP gap, casting doubt on the conclusions of BS18.

In the OU-SALT main sequence (MS) sample, plotted in red in Figure 5.2(b), 31% of hosts have sub-basal activity, compared to only 6% of the BS18 MS (after evolved systems have been removed from the sample). The activity distributions also appear distinct. The BS18 sample contains planetary systems, which are not identifiable from the catalogue. To evaluate the distinctive character of the OU-SALT sample, it is useful to compare it with a sample of stars around which planets have not been detected.

Pace (2013)

Figure 5.3 compares the OU-SALT sample with MS and sub giant (SG) field star samples drawn from Pace (2013) (P13). The samples are limited to $0.4 < B-V < 1.2$ (corresponding to spectral type F2-K5) as $\log(R'_{\rm HK})$ is well calibrated between 0.4 - 1.0 (Isaacson & Fischer, 2010) and unreliable for $B-V > 1.2$ (Noyes et al., 1984). One OU-SALT target falls outside this range: WASP-100 ($B-V = 0.35$). Figure 5.3(a) shows that **31 % of the MS OU-SALT sample has sub-basal activity, compared to only 2 % of P13 MS field stars**. Extrinsic absorption, either by ablated planetary material or the interstellar medium (ISM), must be in operation. In Figure 5.3(b), 75% of the OU-SALT SG population have $\log(R'_{\rm HK}) < -5.1$, compared to 29% of the evolved field-star population. The OU-SALT sample provides the lowest activity values in both MS and SG samples.

To test whether the OU-SALT and P13 distributions are statistically distinct, I employ the *k*-sample Anderson-Darling test. This non-parametric method tests the hypothesis that the populations from which two groups of data were drawn are identical (Stephens, 1974; Scholz & Stephens, 1987). Respective test scores of 31.2 and 32.3 indicate that there are very low probabilities (> 0.001) that the OU-SALT MS and SG planet host populations and corresponding P13 field star samples have been drawn from the same probability distributions. **The activity of planet hosts and field stars is distinct.**

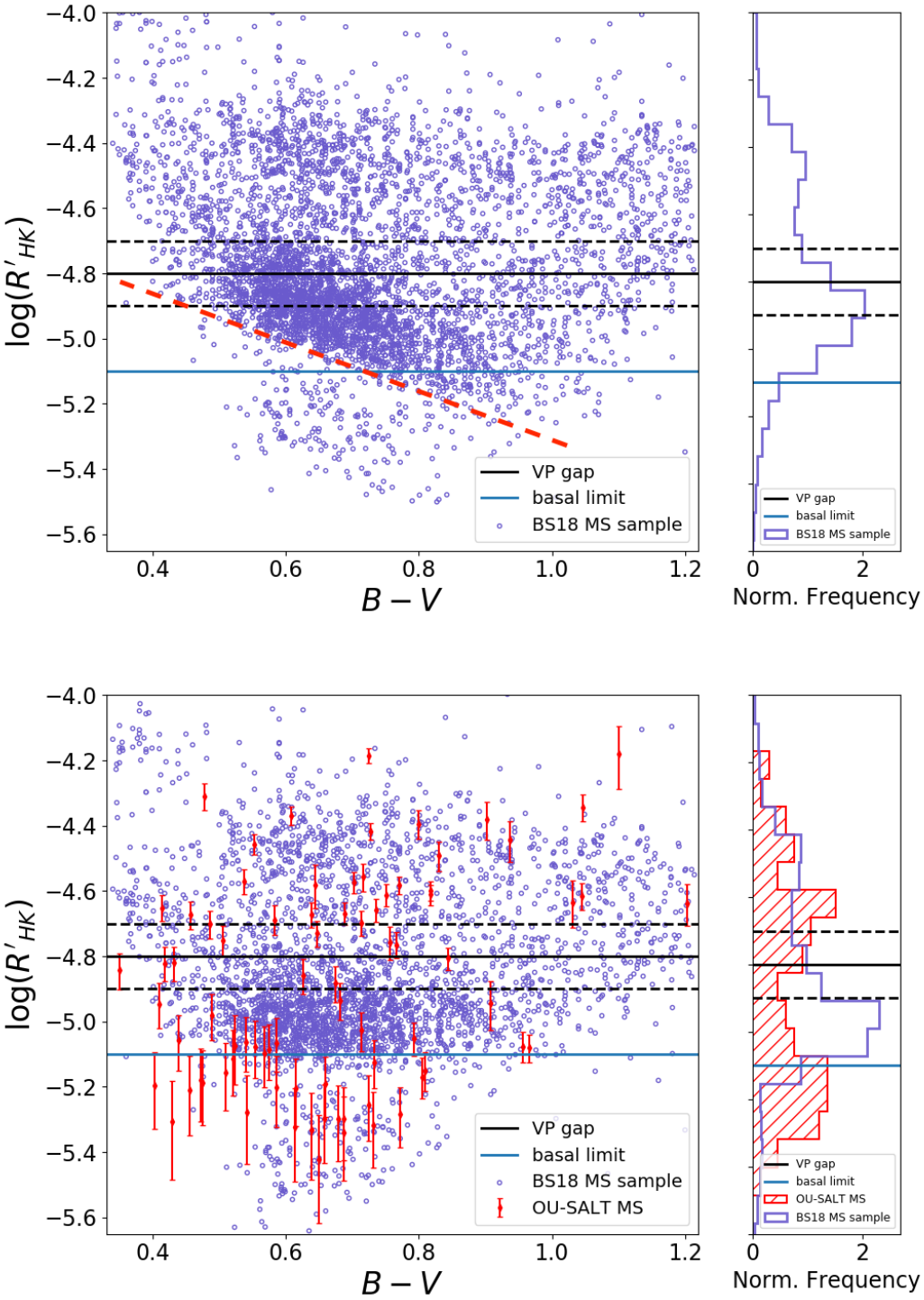

FIGURE 5.2: Top panel: colour-activity plot showing the BS18 MS sample in purple. The histogram in the side panel shows the sample distribution. The position of the Vaughan-Preston gap suggested by Foukal (2018) is marked by dashed black lines, with its mid-point shown by a solid black line. The position of a distinct, boundary-like feature is marked with a dashed red line. Lower plot: corrected BS18 $\log(R'_{\rm HK})$ values. Note the boundary feature now runs along the sub-basal limit. The OU-SALT MS activity values are plotted in red, with error bars that incorporate calibration uncertainty.

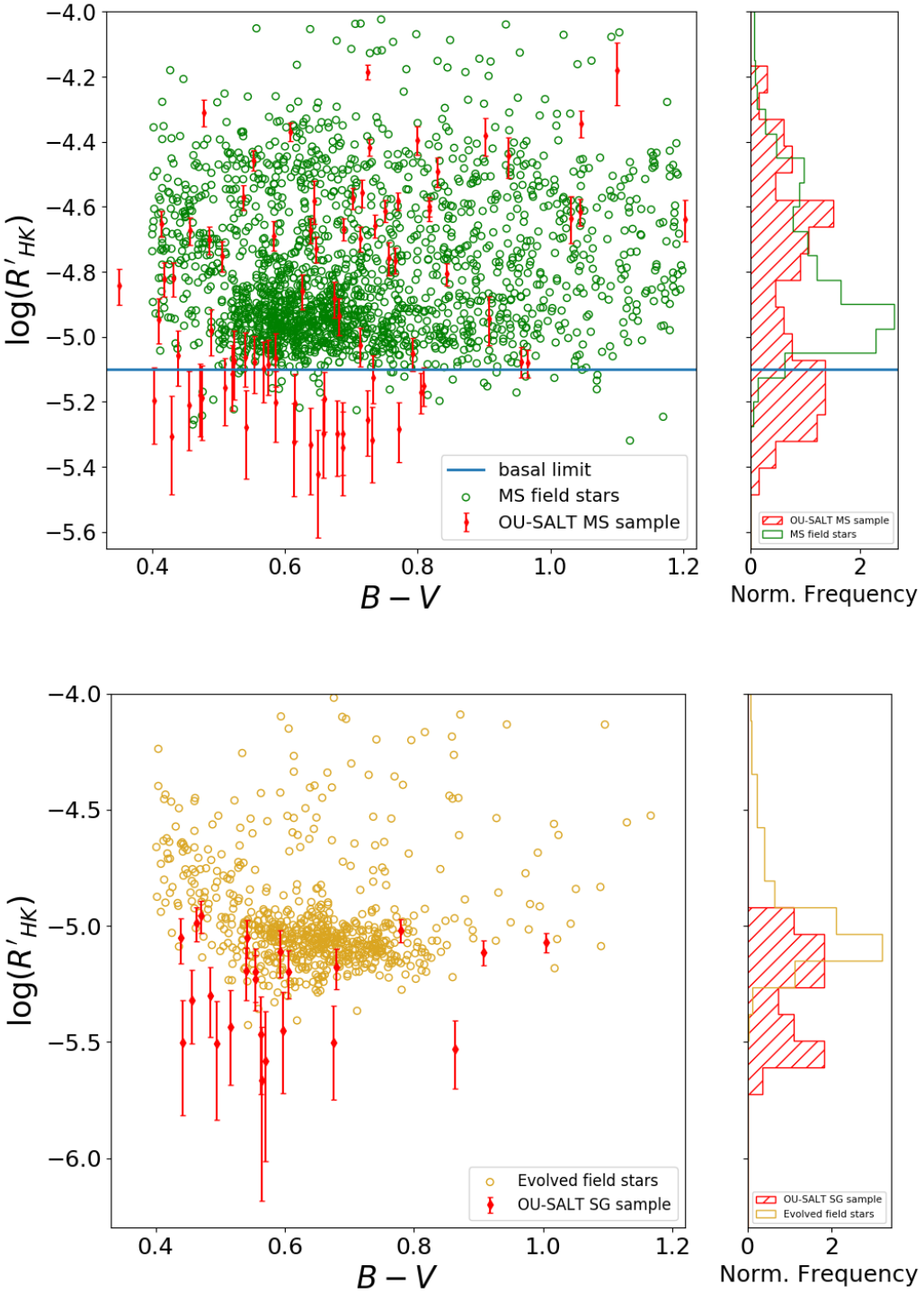

FIGURE 5.3: Top panel: OU-SALT MS activity measurements in context of the $\log(R'_{\rm HK})$ distribution for unevolved MS field stars < 0.45 mag above the empirical MS of Wright (2004). Bottom panel: OU-SALT sub-giants in the context of the $\log(R'_{\rm HK})$ distribution for evolved field stars between 0.45 and 2.0 mag above the Wright (2004) MS. Field stars chosen from P13.

Bimodality

The OU-SALT MS population appears bimodal with peaks at $\log(R'_{\rm HK}) \sim -4.7$ and -5.2. In Hartigans' dip test of unimodality, p-values less than 0.05 indicate significant bimodality and values 0.05-0.10 suggest bimodality with marginal significance (Hartigan & Hartigan, 1985; Freeman & Dale, 2013). A Hartigans' dip test p-value of 0.049 confirms bi-modality in the OU-SALT MS sample. The equivalent p-value for the MS field star P13 population is 0.978. Applying the test to targets with $B-V < 0.8$ only, the dip statistic decreases from 0.57 to 0.53, indicating bimodality is stronger in this sub-sample. The p-value however increases to 0.20, likely due to the smaller sample size.

Bimodality may reflect the presence of the VP gap (see Section 1.3.3): in favourable conditions high activity hosts efficiently strip mass and become enshrouded. Lower activity stars are less obscured due to lower mass loss rates. The VP gap is thus accentuated by mass loss. Taking $\log(R'_{\rm HK}) = -4.9$ to be the dividing line between high and low activity (i.e. the lower boundary of the VP gap suggested by Foukal, 2018), I find the high and low activity sub-samples in the OU-SALT MS to be near equally populous. The MS field star distribution peaks between the high and low activity sub-samples of the OU-SALT MS. The VP gaps widens due to enshrouding. In addition, I propose that the evolution of MS stars from high to low activity may be disrupted by SPI, further accentuating the complexion of the VP gap.

There is possible bimodality in OU-SALT SG population with peaks at ~ -5.1 and ~ -5.5. A Hartigans' dip test p-value of 0.087 suggests bimodality of marginal significance. This value may strengthen for a larger sample. The OU-SALT SG sample has lower average activity than the OU-SALT MS population: median activity is $\log(R'_{\rm HK}) = -5.31$ compared to $\log(R'_{\rm HK}) = -4.63$. In the field star samples, median values are $\log(R'_{\rm HK}) = -4.86$ (MS) and -5.03 (SG). The difference between MS and SG median activity values is 0.51 dex higher in the planet-hosting population. As stars evolve off the MS and expand, planets are exposed to higher radiation levels, making mass-loss, enshrouding and SPI more likely. This may result in a larger spread between median activity values.

5.3 ISM absorption

Sub-basal activity could result from absorption in the ISM. This was found to be a plausible explanation for an apparent lack of activity of WASP-13 (Fossati et al., 2015b) but not likely to be the cause of sub-basal activity of WASP-12 (Fossati et al., 2013). Fossati et al. (2017b) found that absorption for ISM Ca II column densities $N_{\mathrm{CaII}} \gtrsim 12\ \mathrm{cm}^{-2}$ may have a significant influence on the apparent activity of systems where the difference between the radial velocity of the star (ς) and known local ISM clouds (V_{ISM}) is less than $30 - 40$ km s^{-2}. Figure 5.4—which shows $\log N_{\mathrm{CaII}}$ measurements for stars at different distances and lines of sight—indicates that $\log N_{\mathrm{CaII}} \sim 12$ may be expected for Galactic stars more than $100 - 200$ pc away (Welsh et al., 2010). Wyman & Redfield (2013) reach a similar conclusion, finding $\log N_{\mathrm{CaII}} \sim 12$ on average for early-type stars at distances further than 120 pc. As the majority of OU-SALT targets lie beyond 100 pc, it is necessary to investigate bias introduced by ISM absorption.

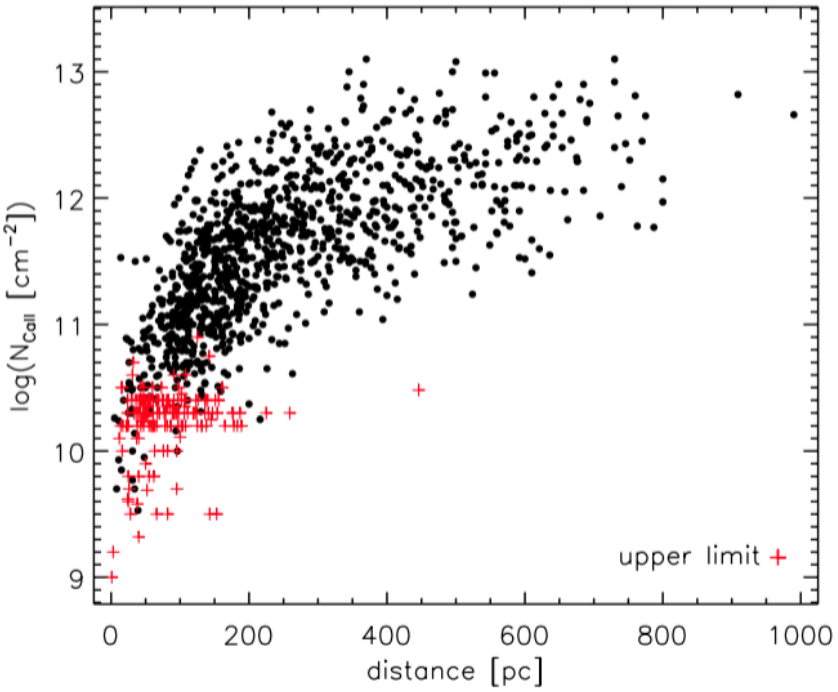

FIGURE 5.4: Measured $\log N_{\mathrm{CaII}}$ as a function of stellar distance for a large sample of stars at different distances and lines of sight Welsh et al. (2010). The red crosses indicate upper limits.

5.3.1 Correcting for absorption

Absorption by the ISM may reduce the S-value if the difference between the velocity of the star through space and the velocity of the intervening ISM cloud is within the $\pm$ 86 km s^{-1} band used to measure the S index, as shown in Figure 5.5(a). ISM absorption is significant when the absorption line is in the region where stellar flux peaks in the core bandpasses (Fossati et al., 2017b). Figure 5.5(b) shows the $S_{\rm MW}$ value decreases with increasing column density.

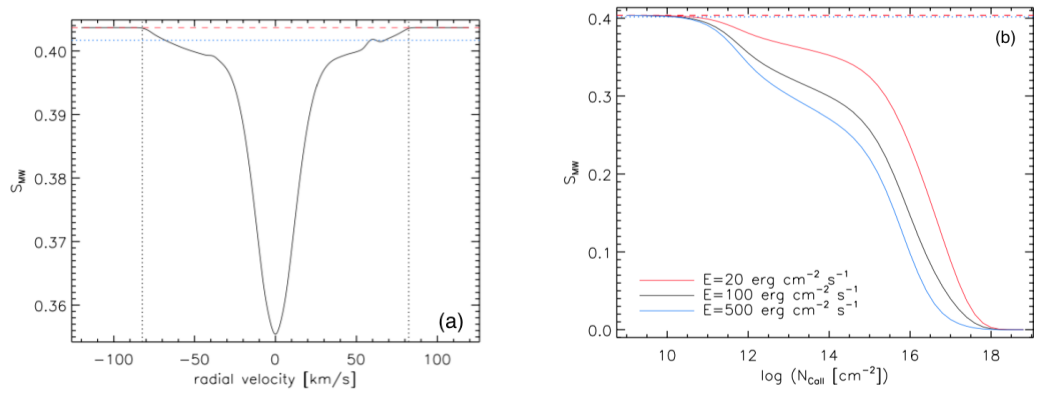

FIGURE 5.5: (a) Variation of $S_{\rm MW}$ with radial velocity difference between system and ISM clouds. (b) Depression of $S_{\rm MW}$ with column density. (Fossati et al., 2017b)

To assess ISM absorption, the relative radial velocities of systems and known local ISM clouds $|\varsigma - \mathrm{ISM}|$ were calculated, with ς values extracted from the Gaia DR2 archive and planet discovery papers. $V_{\rm ISM}$ was calculated using the LISM model[2] (Redfield & Linsky, 2008), which identifies and calculates the radial velocities of ISM clouds crossing the line of sight for each target. ISM contributions to $\log(R'_{\rm HK})$ depression were estimated using a tool[3] developed by Fossati et al. (2017b). Column densities of $\log N_{\rm CaII} = 11.5 - 12$ are assumed (Figure 5.4: Welsh et al., 2010) and average extinction calculated.

[2] http://lism.wesleyan.edu/LISMdynamics.html
[3] http://geco.oeaw.ac.at/software.html

Reduction in observed $\log(R'_{\rm HK})$ due to absorption in the ISM is detailed in Table 5.1. It is assumed in the calculations that all absorption occurs at the lowest $|\varsigma - V_{ISM}|$ value. I therefore provide upper limits for absorption. Measured $\log(R'_{\rm HK})$ values require correction by up to 0.076. Figure 5.6 shows measured and corrected $\log(R'_{\rm HK})$ values for all sub-basal OU-SALT systems. It can be seen that ISM absorption has depressed three targets—WASP-38 (Barros, S. C. C. et al., 2011), WASP-47 (Weiss et al., 2017), and WASP-105 (Anderson, D. R. et al., 2017)—beneath the sub-basal limit but that it is not the driving force behind the majority of unusually inactive stars. Absorption by circumstellar gas clouds remains the other plausible explanation for depressed activity values.

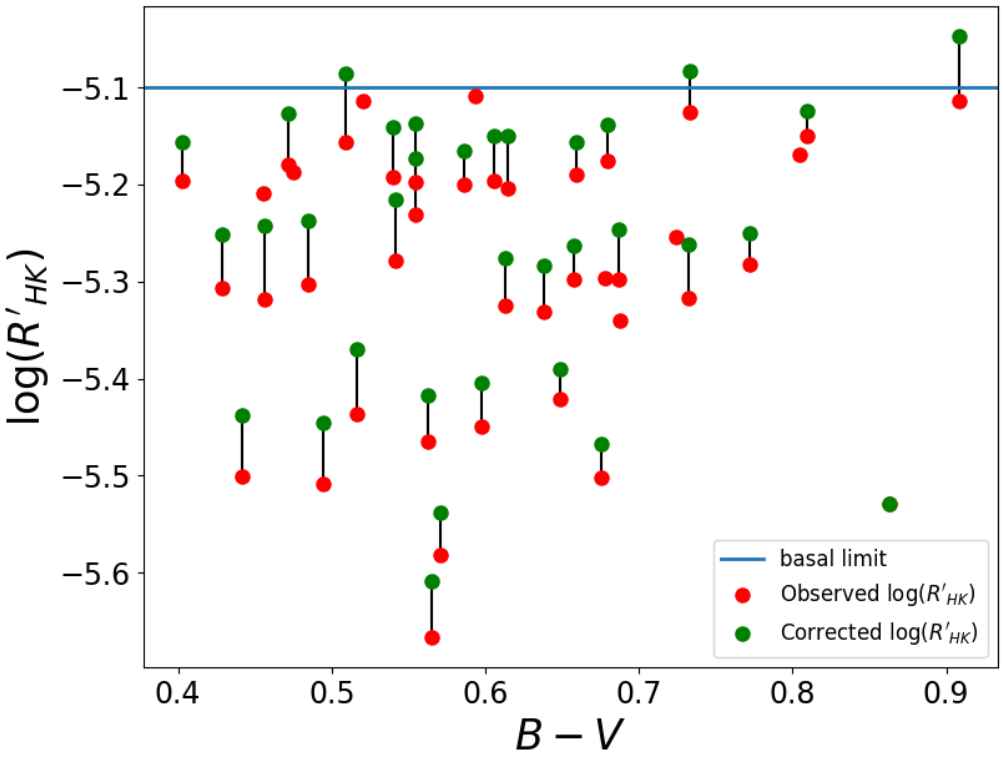

FIGURE 5.6: Observed and corrected $\log(R'_{\rm HK})$ values for OU-SALT targets with $\log(R'_{\rm HK}) < -5.1$. Points without corrections either have no published ς value or are not traversed by LISM clouds (see Table 5.1).

TABLE 5.1: Table detailing stellar radial velocities (ς), the difference between ς and ISM cloud velocities ($|\varsigma - V_{\rm ISM}|$), identity of the clouds that traverse the line-of-sight of each planet host (Redfield & Linsky, 2008), the estimated reduction in $\log(R'_{\rm HK})$ value due to ISM absorption ($\Delta\log(R'_{\rm HK})$), and corrected $\log(R'_{\rm HK})$. Three systems do not have published ς values, while twelve are not traversed by ISM clouds. WASP-100 is too hot for correction.

System	ς [km s^{-1}]	$\lvert\varsigma - V_{\rm ISM}\rvert$ [km s^{-1}]	ISM clouds[a]	$\Delta\log(R'_{\rm HK})$	$\log(R'_{\rm HK})$ (corrected)
CoRoT-7	31.4	10.3,6.9	LIC,Aur	0.060	-4.55
CoRoT-11	40.1	64.2,75.7,72.2	G,Aql,Oph	0.004	-4.30
CoRoT-22	2.2	26.2,37.8,34.3	G,Aql,Oph	0.045	-5.40
CoRoT-28	76.7	101.3,113.8,108.7	G,Aql,Oph	0.000	-5.53
EPIC 219388192	47.3	74.5,71.9	Mic,Aql	0.002	-4.37
HAT-P-27	-18.3	1.3,2.0	Gem,Leo	0.072	-4.53
HATS-2			Gem,Leo,Aur		-4.49
HATS-3			Mic		-5.21
HATS-10	-28.2	2.1,9.1	Mic,Aql	0.049	-5.28
HATS-13	25.8	46.3	Mic	0.009	-4.69
HATS-18	7.7	0.8	Gem	0.070	-4.49
HATS-21	32.7		None		-5.34
HATS-27	-1.1	14.7	G	0.051	-4.62
HATS-29	-19.8		None		-5.29
HATS-30	1.3	3.5,14.4,1.2	LIC,Dor,Vel	0.054	-5.01
HATS-36	-24.4	1.9,3.2	Mic,Aql	0.072	-4.38
HD 73256	29.8		None		-4.42
K2-2	-2.9		None		-4.81
K2-19	7.4	7.5,6.1	Leo,NGP	0.063	-4.55
K2-31	-4.3	24.8	G	0.015	-4.57
K2-32	-1.7	27.8	G	0.022	-5.03
K2-39	24.4	29.3	LIC	0.014	-5.06
K2-96	19.3		None		-5.17
K2-106	-14.2		None		-5.25
K2-135	32.1		None		-4.64
KELT-10	31.6	48.8	Aql	0.034	-5.17
KELT-11	35.2	19.2,31.2	Gem,Leo	0.032	-4.99
KELT-15	12.3	4.8	Blue	0.058	-5.61
Qatar 2	-23.6	10.5,4.3	Leo,NGP	0.072	-4.27
WASP-5	20.7	21.7	LIC	0.024	-4.65
WASP-6	12.5	15.1,25.5	LIC,Cet	0.034	-4.62
WASP-7	-29.4	28.1,4.0	Aql,Vel	0.072	-4.75
WASP-8	-1.9	1.9	LIC	0.068	-4.51
WASP-15	-1.4	10.5,20.5	Gem,NGP	0.055	-5.25
WASP-17	-48.0	19.8	G	0.040	-5.16
WASP-18	4.1	0.7,26.0,7.1	LIC,Dor,Cet	0.053	-5.13
WASP-25	-1.6	3.2,8.3,13.6	Gem,Leo,NGP	0.065	-4.66

System	ς [km s^{-1}]	$\lvert\varsigma - V_{ISM}\rvert$ [km s^{-1}]	ISM clouds[a]	$\Delta \log(R'_{HK})$	$\log(R'_{HK})$ [corrected]
WASP-26	9.6	5.1,10.9,5.4	LIC,Mic,Cet	0.060	-5.14
WASP-28	25.8		None		-5.12
WASP-32	18.4		None		-5.07
WASP-34	50.0	37.8,49.1,43.9	Gem,Leo,Aur	0.033	-5.16
WASP-36	-11.3	24.0	LIC	0.023	-4.67
WASP-38	-9.1	11.7,14.0,14.9	LIC,Mic,Oph	0.071	-5.09
WASP-43	-3.7	23.3,10.2,13.8	Gem,Leo,Aur	0.051	-4.13
WASP-44	-4.0	8.7,8.2	LIC,Cet	0.056	-4.70
WASP-45	5.0	3.4	LIC	0.068	-4.88
WASP-46	-2.8	14.6	Vel	0.036	-4.55
WASP-47	-25.9	17.9	LIC	0.042	-5.08
WASP-50	25.2	7.9	LIC	0.059	-4.13
WASP-51	45.4	28.7,25.5	LIC,Aur	0.049	-4.93
WASP-52	0.5		None		-4.38
WASP-55	-3.1	15.5	NGP	0.059	-5.02
WASP-62	15.5	9.7,8.7,1.2,0.0	G,Blue,Cet,Vel	0.067	-4.68
WASP-63	-23.1	34.1	Blue	0.034	-5.47
WASP-65	-3.2	17.2,33.2	LIC,Gem	0.035	-4.85
WASP-67	0.3	26.4,18.6	Mic,Aql	0.028	-4.74
WASP-69	-9.4		None		-4.62
WASP-70	-64.4	42.6	Mic	0.031	-5.39
WASP-72	37.4	24.9,19.3	LIC,G	0.065	-5.24
WASP-73	11.1	27.4	Vel	0.048	-5.42
WASP-74	-15.3	1.7	Eri	0.046	-5.15
WASP-75	2.7	6.2	LIC	0.064	-5.22
WASP-76	-1.1	15.6	LIC	0.067	-5.37
WASP-77	4.1	12.3	LIC	0.045	-4.62
WASP-78	1.3	16.8	LIC	0.062	-5.45
WASP-79	5.2	10.8,6.8,29.3	LIC,Blue,Dor	0.065	-4.76
WASP-80	9.8	27.3	Eri	0.006	-4.44
WASP-82	-23.3	46.3,35.8	LIC,Hyades	0.063	-5.44
WASP-87	-13.6	0.8	G	0.075	-4.87
WASP-88	-5.2	18.4	Vel	0.070	-4.98
WASP-89	21.4	43.5	Mic	0.003	-4.44
WASP-90	4.4	17.4	Eri	0.076	-5.24
WASP-91	3.0	5.4	Vel	0.067	-5.01
WASP-94A	-8.3	12.0	Mic	0.064	-4.99
WASP-94B	-8.4	11.9	Mic	0.063	-5.02
WASP-95	6.4	14.2	LIC	0.046	-4.81
WASP-96	-0.9	0.5,24.2	LIC,Dor	0.056	-5.26
WASP-97	7.2	14.9,0.7,4.1	Dor,Cet,Vel	0.051	-5.25
WASP-99	25.1	14.6,2.1	G,Dor	0.052	-5.14
WASP-100	29.4	23.1,14.1,13.9	G,Cet,Vel		-4.84

System	ς [km s^{-1}]	$\lvert\varsigma - V_{ISM}\rvert$ [km s^{-1}]	ISM clouds[a]	$\Delta \log(R'_{HK})$	$\log(R'_{HK})$ [corrected]
WASP-101	42.9	30.4	Blue	0.028	-4.67
WASP-103	-40.3	18.6	LIC	0.032	-4.54
WASP-104	28.5	22.3	Leo	0.033	-5.25
WASP-105	25.3	21.8,1.0,16.5,11.8	LIC,Dor,Cet,Vel	0.066	-5.05
WASP-108	46.4	62.4	G	0.019	-5.08
WASP-109	-15.5	12.6,8.9	G,Gem	0.070	-4.58
WASP-111	-19.3	9.7,2.4	LIC,Mic	0.061	-4.93
WASP-114	0.1		None		-5.11
WASP-117	-16.0	21.9,43.2,32.0	LIC,Dor,Cet	0.049	-5.01
WASP-118	7.5	4.7	LIC	0.068	-4.89
WASP-119	9.2	3.9,2.6,4.1	G,Cet,Vel	0.062	-4.87
WASP-120	19.4	4.7,9.6,7.1	G,Blue,Dor	0.076	-4.98
WASP-122	34.6	25.0	Blue	0.038	-5.14
WASP-123	17.3	37.1	Aql	0.034	-5.26
WASP-126	29.5	25.9,20.0,18.5	G,Cet,Vel	0.047	-5.28
WASP-127	-8.2	25.4,13.3	Gem,Leo	0.054	-5.15
WASP-130	1.6	20.8	NGP	0.035	-4.99
WASP-131	-19.4	6.8,3.0	Gem,NGP	0.057	-5.17
WASP-132	31.6	53.3	G	0.002	-4.63
WASP-138			LIC		-5.19
WASP-139	-11.8	27.3,44.5	G,Dor	0.025	-5.12
WASP-140	2.0	16.2	LIC	0.030	-4.36
WASP-142	48.6	42.3	G	0.043	-5.54
WASP-156	9.3	7.1	LIC	0.063	-5.02

5.4 Planet parameters

The relationship between chromospheric activity and planetary parameters will be explored in this section. Table 5.2 details the planetary parameters for each target.

5.4.1 Surface gravity

Surface gravity may be derived empirically with parameters measured by radial velocity (RV) and transit studies:

$$g_{\rm P} = \frac{2\pi}{P_{\rm orb}} \frac{(1-e^2)^{1/2} K_\star}{(R_{\rm P}/a)^2 \sin i}. \tag{5.1}$$

The eccentricity (e) and RV semi-amplitude ($K_\star$) are obtained from the host's reflex RV curve. The ratio of the planet radius to semi-major amplitude ($R_{\rm P}/a$), along with the inclination (i), are obtained from the transit light curve (Perryman, 2018). The orbital period $P_{\rm orb}$ may be determined from either the RV or transit curve.

The correlation between ($g_{\rm P}$) and activity is well documented. Knutson et al. (2010) found a correlation between activity and the emission (i.e. dayside) spectra of transiting planets: high gravity planets with non-inverted temperature models were found around high-activity stars, while planets with temperature inversions were found around lower activity stars. Hartman (2010) found a correlation between chromospheric emission and planetary surface gravity. They identify low emission of stars with planets that have low surface gravity, and find that planets with strong surface gravity tend to be found around more active stars. Figueira et al. (2014) confirmed this correlation. Invoking the rational of Haswell et al. (2012), Lanza (2014) explained this correlation using a model that assumes plasma composed of planetary material evaporated under the action of EUV radiation forms clouds that absorb Ca II H & K cores, providing a theoretical relationship between $\log(R'_{\rm HK})$ and $g_{\rm P}^{-1}$ (Lammer et al., 2003; Lecavelier des Etangs et al., 2004; Lecavelier des Etangs, 2006; Sanz-Forcada et al., 2010, 2011).

Planets with lower gravity are more prone to mass loss. For example, Yan & Guo (2019) find Lyman α absorption by planetary atmospheres is large where stellar irradiation is high and mean density low. Mass loss from HJs is lower due to large gravitational potential. Faster evaporation rates lead to higher column density of absorbing circumstellar material, and lower levels of observed emission.

Fossati et al. (2015b) use a cluster-weighted model (CWM) to find two distinct distributions in the $\log(R'_{\rm HK}) - g_p$ correlation of a sample of HJs (with $4200 < T_{\rm eff} < 6200$ K, $a < 0.1$ AU, $M_{\rm P} > 0.1 M_{\rm J}$, and $V < 13$). Figure 5.7 shows there is not a strict activity division between the distributions but instead each distribution includes systems that share similar physical qualities. The authors argue that the two component distribution is a consequence of the Vaughan-Preston (VP) gap: planets orbiting high-activity stars experience higher levels of EUV flux, meaning they are more prone to mass loss than planets orbiting lower activity stars. The $\log(R'_{\rm HK}) - g_{\rm P}$ correlation is stronger in the high activity population than the low activity population.

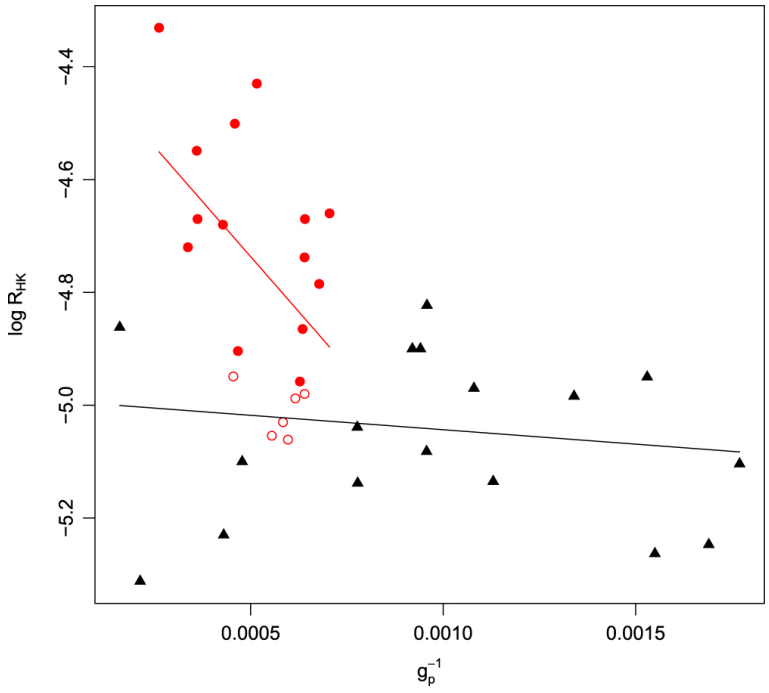

FIGURE 5.7: Inverse of planetary gravity (in cm^{-1} s^2) vs. activity with two best-fit regression lines from a cluster-weighted model in black and red (Fossati et al., 2015a). The points assigned to each regression have the same colour code. The open circles are assigned to the high activity group but with a posteriori probability of $0.5 - 0.65$, may also belong to the low activity group. The horizontal extent of the plotted linear fits indicates the x-axis range, which is also the case for subsequent correlation plots.

Activity–surface gravity relationship in the OU-SALT sample

Surface gravity values have been published for 93 OU-SALT targets. The R_P/a ratio is unavailable for the two non-transiting hosts: HD 73256 (Udry et al., 2003) and WASP-94B (Neveu-VanMalle et al., 2014), and so g_P can not be calculated for these systems. For the remaining nine targets g_P has been calculated using Equation 5.1. Two systems—K2-32 (Heller et al., 2019) & K2-39 (Petigura et al., 2017a)—do not have published inclination i. This has been assumed to be 90°. No errors are available for the eccentricity of K2-96 (Vanderburg et al., 2016). Figure 5.8 shows the inverse of surface gravity plotted against activity for OU-SALT MS and SG populations. The inset boxes detail:

- γ is the slope of plotted least squares linear regression;
- ρ is the Spearman rank correlation coefficient—a non-parametric measure of the monotonic relationship between variables where a perfect correlation is displayed by a value of 1 and a perfect anticorrelation by -1 (Spearman, 1904). This is the most widely cited measurement of correlation. The associated p-value is the probability that the observed correlation is due to chance. Its value is between 0 and 1: values less than 0.01 provide very strong evidence to reject the null hypothesis; $0.01 - 0.05$ provide strong evidence, $0.05 - 0.10$ provide weak evidence; and values exceeding 0.1 provide very weak evidence;
- τ is the Kendall tau rank correlation coefficient which, like ρ, is non-parametric and provides scores between -1 and 1 (Kendall, 1938). However, calculations are based on concordant and discordant pairs rather than deviations, and are much less sensitive to error and discrepancies than ρ. Associated p-values are more accurate for small sample sizes (Xu et al., 2013).

I generally cite ρ in my analysis for ease of comparison with other studies but τ is the more informative correlation coefficient. Each of these statistics are presented with associated uncertainties (which is unusual for ρ and τ) calculated with the perturbation technique (Curran, 2014). This involves a Monte Carlo-type approach of creating multiple new data sets where x (a planetary or stellar variable, or interaction proxy) and y ($\log(R'_{HK})$) are perturbed from their measured value by a number

drawn randomly from a Gaussian distribution of standard deviation equal to the measured error (Δx and Δy respectively). This may be expressed as $x = x + \mathcal{G}(\Delta x)$ and $y = y + \mathcal{G}(\Delta y)$. The above-detailed statistics for each new data set are collected into arrays. The root-mean-square of the each statistic's distribution provides 1 σ uncertainty values.

There is a strong negative $\log(R'_{\rm HK}) - g_{\rm P}^{-1}$ correlation in the high activity MS sample ($\rho = -0.41 \pm 0.02$), with a correspondingly small p-value ($p = 0.01 \pm 0.006$). The anti-correlation is less strong in the low-activity MS sample ($\rho = -0.31 \pm 0.03$, $p = 0.06 \pm 0.026$), although marginally strengthens to $\rho = -0.33 \pm 0.04$ ($p = 0.06 \pm 0.02$) when the three lowest $g_{\rm P}^{-1}$ planets are removed from the distribution. These results confirm the findings of Hartman (2010), Figueira et al. (2014) and Fossati et al. (2015b). Anti-correlations are expected as planets with lower surface gravity more readily lose their atmospheres. Mass-loss and enshrouding is less likely to occur in high $g_{\rm P}$ systems. The $\log(R'_{\rm HK}) - g_{\rm P}^{-1}$ anti-correlation is stronger in the high activity population as planets orbiting higher activity stars are more highly irradiated. There is a weak, non-significant anti-correlation in the SG sample ($p = 0.81 \pm 0.142$). The difference in slope ($\Delta\gamma = \gamma_{\rm low} - \gamma_{\rm high} = 0.82 \pm 0.15$) for the OU-SALT MS is less marked than reported in Fossati et al. (2015b)($\Delta\gamma = 6.91$). The OU-SALT dataset is larger, homogeneous and samples more $g_{\rm P}$ parameter space than Fossati et al. (2015b), and thus provides a more informative result. If more low $g_{\rm P}$ systems with high activity are identified, $\Delta\gamma$ may decrease further.

Exceptions to the general trends in the MS population are apparent from Figure 5.8, where systems with some of the highest $g_{\rm P}$ values have sub-basal activity. The colour scale shows that these tend to be systems where $T_{\rm eq}$ is high ($\gtrsim 1800K$), i.e. systems approaching the UHJ regime where there is sufficient heat in the planetary atmosphere to overcome the energy threshold required to escape the strong gravitational pull. This interplay between $T_{\rm eq}$ and $g_{\rm P}$—leading to either atmospheric retention or loss—causes a bifurcation in the MS $\log(R'_{\rm HK}) - g_{\rm P}^{-1}$ distribution for the highest $g_{\rm P}$ systems ($g_{\rm P} \leq 0.05$ m^{-1} s^2).

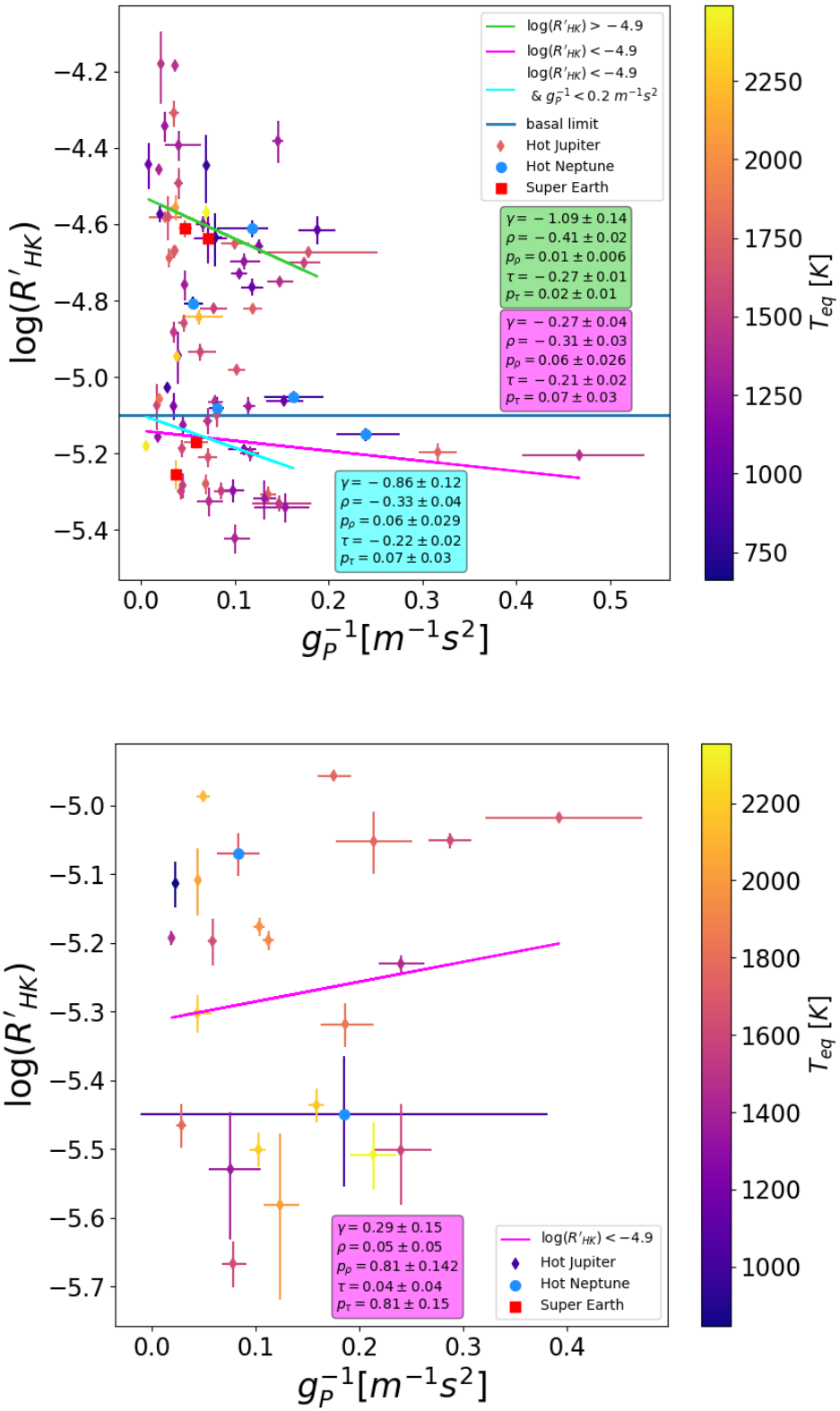

FIGURE 5.8: Top panel: the inverse of planetary surface gravity plotted against activity for the OU-SALT MS. The colour scale shows $T_{\rm eq}$. Bottom panel: as per the top panel but for the OU-SALT SG sub-sample. The two RV-discovered targets are not plotted. See text for description of the in-box statistics. (Subsequent correlation plots also adopt this panel arrangement.)

5.4.2 Equilibrium temperature

Equilibrium temperature ($T_{\rm eq}$) is the theoretical temperature of a planetary black body being heated only by its parent star. Planets with high $T_{\rm eq}$ are susceptible to complete atmospheric loss (Gupta & Schlichting, 2019). $T_{\rm eq}$ may be calculated:

$$T_{\rm eq} = T_{\rm eff}\sqrt{\frac{R_\star}{a}}[f(1-A_B)]^{\frac{1}{4}}, \tag{5.2}$$

where f is the re-radiation factor and A_B is the Bond albedo (the "total radiation reflected from an object compared to the total incident radiation": Haswell, 2010). Published values of $T_{\rm eq}$ are inhomogeneous as different albedo and heat redistribution parameters are adopted by different authors. For example, Charbonneau et al. (2005) use $f = 1$ where a planet is isotropic (fast rotators) and $f = 2$ if only the planet day side re-radiates absorbed energy, such as for tidally locked planets with no atmospheres or oceans. López-Morales & Seager (2007) adopt $f = 1/4$ for efficient heat distribution and isotropic re-radiation over the whole planet, and $f = 2/3$ for instantaneous re-radiation of incident radiation with no heat redistribution.

Hot Jupiters have a range of albedos depending on temperature, which influences the formation of clouds (Sudarsky et al., 2000). The formation of atmospheric features such as sulphur hazes can drastically alter albedo (Gao et al., 2017). Most HJs have been found to host clouds and have very low albedo (Parmentier et al., 2016). For example, Kipping & Spiegel (2011) found the geometric albedo ($A_{\rm g}$)—which is the "amount of radiation relative to that from a flat Lambertian surface which is an ideal reflector at all wavelengths" (De Pater & Lissauer, 2015)—of TrES-2b to be 0.0253 ± 0.0072. Other low albedo HJs include HAT-P-7b ($A_{\rm g} \leq 0.03$: Morris et al., 2013), Kepler-12b ($A_g = 0.14 \pm 0.04$: Fortney et al., 2011), and Kepler-423b ($A_g = 0.055 \pm 0.028$: Gandolfi et al., 2015). However, some hot Jupiters have been found to be brighter, such as Kepler-7b with $A_{\rm g} \simeq 0.35$ (Demory et al., 2011, 2013). Albedos of close-in smaller planets have also found to be low. Sheets & Deming (2017) find super-Earths, mini-Neptunes and Super-Neptunes have albedos of $0.11 \pm 0.06, 0.05 \pm 0.04$ and 0.11 ± 0.08 respectively.

Activity–equilibrium temperature relationship in the OU-SALT sample

OU-SALT host equilibrium temperatures have been calculated using Equation 5.2 to explore f and A_B parameter space with $f = 1/4, 2/3, 1.0$ & 2.0, and $A_B = 0.0, 0.1, 0.3$ & 0.5. Figure 5.9 shows that a higher Bond albedo produces a lower T_{eq} with ~ 50 K decreases in T_{eq} for each 0.1 increase in albedo. Changes in f have a more significant effect, with a ~ 1500 K increase between $f = 1/4$ and $f = 2$. I adopt $f = 1/4$ and $A_B = 0$ as the majority of close-in planets are dark, tidally-locked, and have cloudy atmospheres that transfer heat to night-sides (Parmentier et al., 2016). The T_{eq} value for each system is listed in Table 5.2. Uncertainties, which are propagated from T_{eff}, $R_\star$, and a, range from $15 - 171$ K, with a median of 38 K in the MS sample and 54 in SG sample.

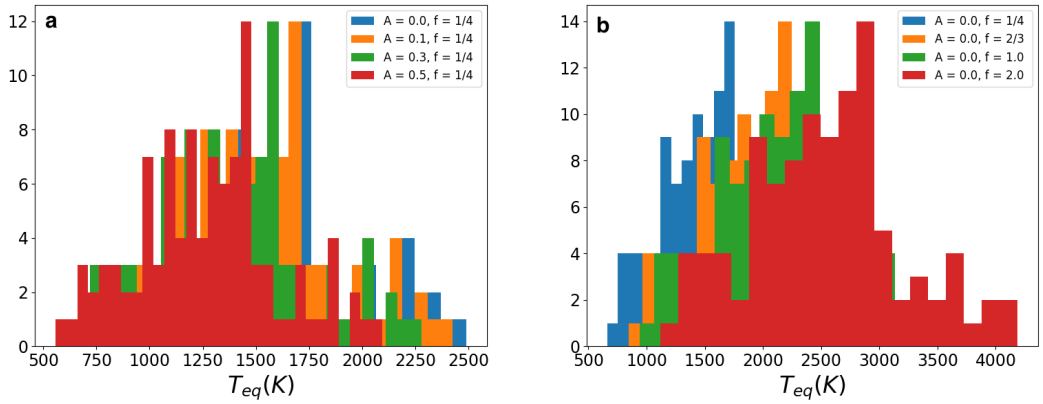

FIGURE 5.9: Frequency histograms showing how T_{eq} varies with A_B (left) and f (right), calculated using Equation 5.2. Values trialled are $f = 1/4, 2/3, 1.0$ & 2.0 and $A_B = 0.0, 0.1, 0.3$ & 0.5.

Figure 5.10 shows a weak $\log(R'_{HK}) - \log T_{eq}$ anti-correlation in high and low activity MS samples with $\rho = -0.006 \pm 0.04$ ($p = 0.68 \pm 0.16$) and $\rho = -0.12 \pm 0.01$ ($p = 0.47 \pm 0.05$) respectively. The p-values provide only very weak evidence for rejecting the null hypothesis. The correlation is similarly weak in the SG population. The linear fit for the whole MS sample shown in blue does not differ significantly to the high- and low-activity sub-samples. A bifurcation in the MS sample is apparent where $\log(T_{eq}) < 3.25$ (1780 K). The horizontal dashed lines in Figure 5.10 mark the start of the UHJ temperature regime ($T_{eq} > 2200$ K: Bell & Cowan, 2018), where dissociation and improved heat transport to the night-side may lead to greater mass

loss and, hence, enshrouding. Ten OU-SALT targets are UHJs, 50% of which are sub-basal in the MS population.

Surface gravity is more strongly correlated to activity in the OU-SALT population than $T_{\rm eq}$. However, $T_{\rm eq}$ has been calculated with an arbitrary choice of f and A_B values, which in reality may be quite different. A weaker $\log(R'_{\rm HK}) - \log T_{\rm eq}$ correlation may result from the method of calculation. It may also be because SPI and enshrouding effects both increase with $T_{\rm eq}$ but act in opposite senses on observed $\log(R'_{\rm HK})$. $T_{\rm eq}$ may play an important role in borderline cases between mass-loss and atmospheric retention. Its role may also become more apparent if further-out, cooler planets are sampled.

5.4.3 Semi-major axis

Planets orbiting close to their host star are heavily irradiated and have strong star-planet interactions, making semi-major axis a key parameter in activity studies. The semi-major axis may be estimated from the stellar mass and orbital period. The inverse of semi-major axis is commonly used as a proxy for SPMI (e.g. Poppenhaeger et al., 2010; Miller et al., 2015; Staab et al., 2017).

The relationship between activity and semi-major axis has been the subject of several studies. Kashyap et al. (2008) found that stars with planets closer than 0.15 AU had four times higher X-ray luminosity than those with $a > 1.5$ AU. Poppenhaeger & Schmitt (2011) subsequently found these trends to be traced back to selection effects and observational biases. Poppenhaeger et al. (2010) found no significant correlations between activity indicator $L_{\rm X}/L_{\rm bol}$ with semi-major axis in the same sample. Martins et al. (2011) found no significant correlation between $\log(R'_{\rm HK})$ and a in either a sample of 19 stars from Kashyap et al. (2008) ($\rho = -0.26 \pm 0.06, p = 0.30 \pm 0.12$) or a sample of 13 stars from Poppenhaeger et al. (2010) ($\rho = -0.29 \pm 0.08, p = 0.34 \pm 0.10$). Both samples contained targets with X-ray luminosity and $a < 0.2$ AU. In a sample of 210 planet hosts, Krejčová & Budaj (2012) use the Kolmogorov-Smirnov and Student's t-test to verify that cool stars ($T_{\rm eff} < 5500$ K) hosting planets

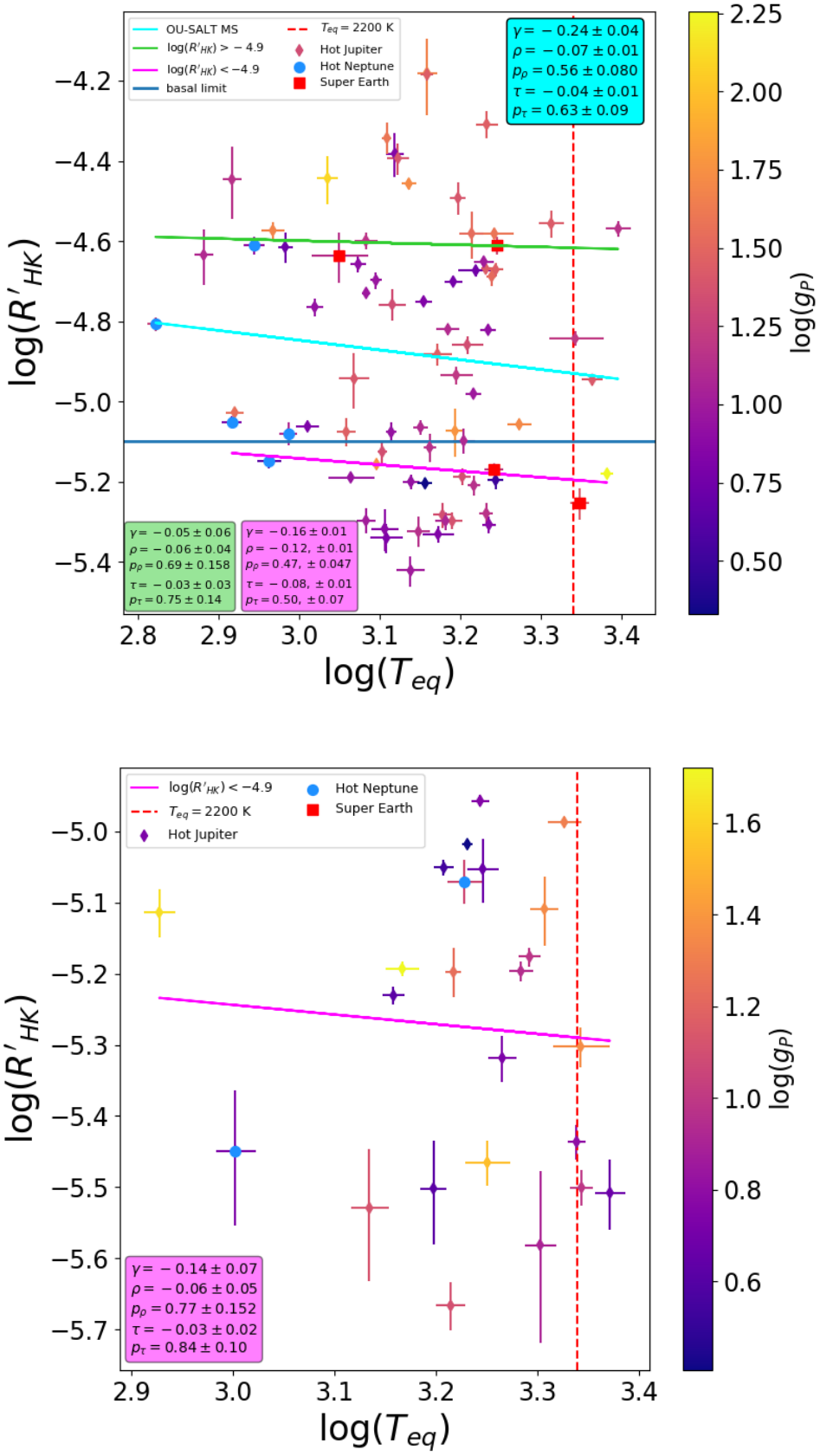

FIGURE 5.10: Equilibrium temperature plotted against activity for MS (top) and SG (bottom) populations, with a $\log g_P$ colour scale. T_{eq} has been calculated homogeneously with $f = 1/4$ and $A_B = 0$. Bifurcation is apparent in the cooler MS population. Activity drops in the hotter population.

with $a < 0.15$ AU have higher $\log(R'_{\rm HK})$ values than stars with more distant planets. They also find close-in planet hosts display a wider range of $\log(R'_{\rm HK})$ values, mainly caused by hotter stars with transiting planets. When looking at cold RV-detected stars only, this trend becomes less pronounced and there is no $\log(R'_{\rm HK}) - a$ correlation.

The ratio of the planet mass to semi-major axis squared ($M_{\rm P}/a^2$) is a commonly-used SPMI proxy. Miller et al. (2015) found $L_{\rm X} - a$ and $L_{\rm X}/L_{\rm bol} - a$ correlations in a sample of 198 FGK stars that was driven by "extreme" systems with $M_{\rm P}/a^2 > 450$ M$_{\rm J}$ AU^{-2}. The median $M_{\rm P}/a^2$ for transiting planets recorded in the TEPCAT database with both $M_{\rm P}$ and a values is ~ 220 M$_{\rm J}$ AU^{-2}, with around 1/3 having $M_{\rm P}/a^2 > 450$ M$_{\rm J}$ AU^{-2}. In a sample of transiting hosts drawn from Figueira et al. (2014) with $4200 \leq T_{\rm eff} \leq 6200$, $V \leq 13$ mag, $E(B-V) \leq 0.06$mag, $a \leq 0.1$ AU and $M_{\rm P} \geq 0.1 M_{\rm J}$, Fossati et al. (2015a) found $\log(R'_{\rm HK}) - a$ correlations of $\rho = -0.46$ in the whole sample, and $\rho = -0.60$ for targets within 0.06 AU.

Activity–semi-major axis relationship in the OU-SALT sample

Figure 5.11 shows the activity–semi-major axis distribution. There is a strong anti-correlation in the MS sequence with $\rho = -0.29 \pm 0.01$ ($p = 0.01 \pm 0.003$). This is comparable to trends reported by Martins et al. (2011) noted above except that the p-value indicates a statistically significant correlation in the OU-SALT MS population. The correlation is less strong than those reported by Fossati et al. (2015b). A stronger $\log(R'_{\rm HK}) - a$ correlation is present for targets within $a < 0.05$ AU ($\rho = -0.43 \pm 0.01$, $p = 1.65 \times 10^{-3} \pm 3.56 \times 10^{-4}$). Figure 5.11 also shows the $\log(R'_{\rm HK}) - a$ distributions for high- and low-activity MS sub-samples. In the high activity MS sample, activity drops as a increases ($\rho = -0.42 \pm 0.01$, $p = 0.01 \pm 0.003$). This suggests SPI has a more significant effect on measured activity than mass loss in this population. In the SG population, there is a weak positive correlation ($\rho = 0.14 \pm 0.02$, $p = 0.32 \pm 0.03$) for the whole sample but a strong negative correlation where $a < 0.05$ AU ($\rho = -0.33 \pm 0.02$, 0.52 ± 0.061). Around 25% of the SG sub-basal population has $a > 0.06$ AU. These large p-values provide weak evidence for rejecting the null hypothesis.

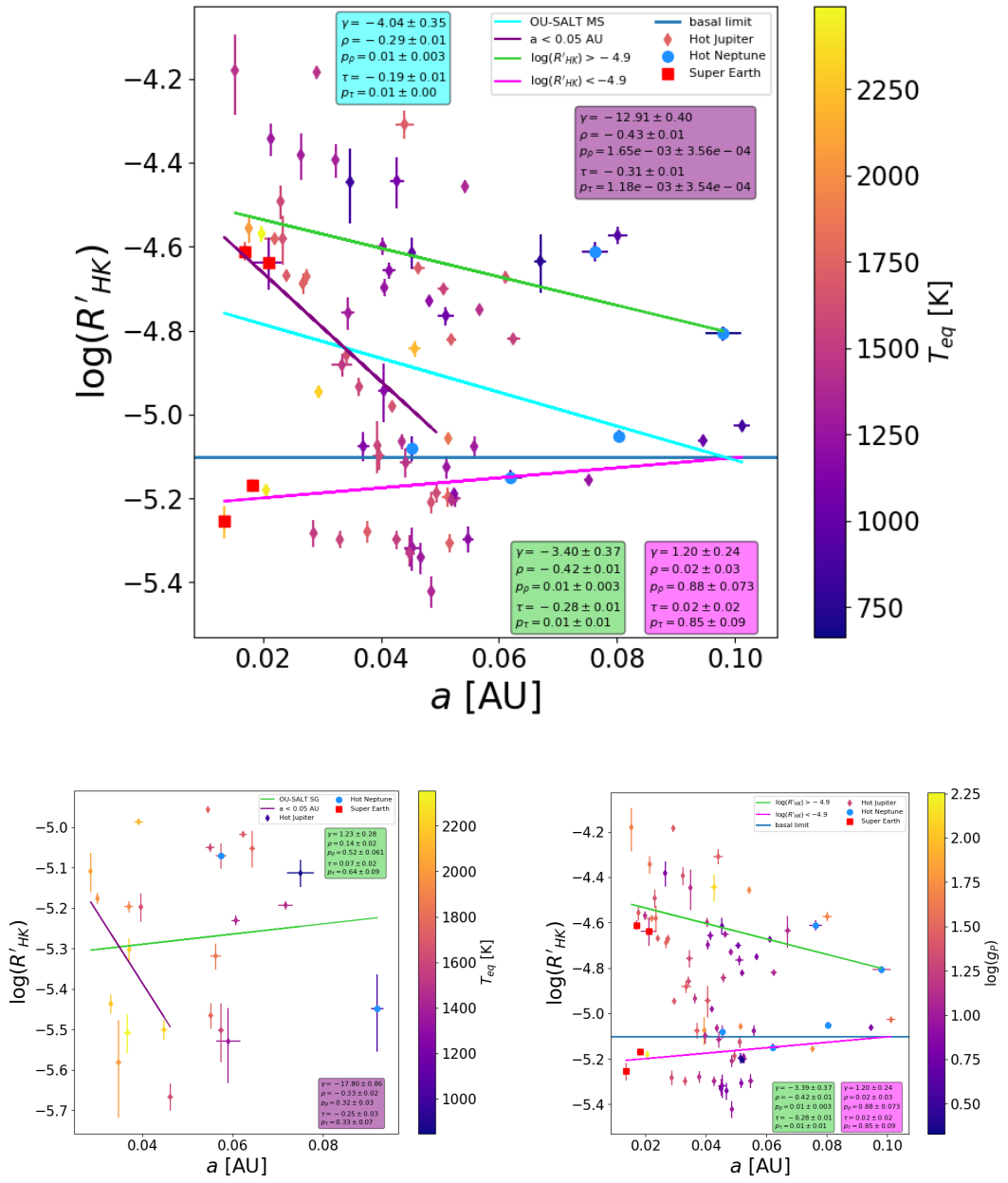

FIGURE 5.11: Semi-major axis plotted against activity for: OU-SALT MS (top) and SG (bottom left) samples with a colour scale showing equilibrium temperature; and for the MS (bottom right) with a colour scale showing the logarithm of surface gravity.

For systems with $a < 0.05$ AU, semi-major axis appears to have a marked effect on observed stellar activity: larger star-planet separation leads to lower activity. At close separations, SPMI and SPTI may cause an increase in the activity of the star—an effect that decreases with distance. Higher irradiation at close separations may lead to increased enshrouding. The colour scales in Figure 5.11 show that the closest planets generally have the highest equilibrium temperature and surface gravity. For systems with $a < 0.04$ AU, there is a clear a bifurcation between the high and low activity populations. Where these closest-in systems have moderate $T_{\rm eq}$, they retain their atmospheres, making intrinsic activity (which may have been enhanced through SPI) observable. However, where $T_{\rm eq} \gtrsim 1780$ K, activity tends to fall below the basal limit. These systems are approaching the UHJ H_2 dissociation regime where they may more readily lose their atmospheres, which cloaks activity signals from the host star. Lower $g_{\rm P}$ systems beyond 0.04 AU appear to more readily lose their atmospheres than their higher $g_{\rm P}$ counterparts, and exhibit sub-basal activity. The sub-basal systems predominantly have $\log g_{\rm P} \lesssim 1.5$. Around 89 % of sub-basal MS targets have a < 0.06 AU, suggesting mass loss occurs mainly in the closest-in systems in accordance with simple expectations.

5.4.4 Mass

Massive planets have strong tidal interactions with their hosts and large $g_{\rm P}$ values. Scharf (2010) found a correlation between $M_{\rm P}$ and host X-ray luminosity for systems with $a < 0.15$ AU observed by the ROSAT All-Sky Survey (Voges et al., 1999). Poppenhaeger & Schmitt (2011) also identified a $L_{\rm X} - M_{\rm P}$ correlation in this data, and in a more complete sample incorporating data from *XMM-Newton* (Lumb, 2012) but found these correlations had their origins in the correlation between stellar distance to the target d and $M_{\rm P}$ i.e. planet detectability by the RV method. They note that, in an unbiased sample, activity and $M_{\rm P}$ must not depend on d. Martins et al. (2011) found $\log(R'_{\rm HK}) - M_{\rm P}$ correlations of $\rho = 0.13 \pm 0.07$ ($p = 0.61 \pm 0.19$) in the Kashyap et al. (2008) sample and of $\rho = 0.20 \pm 0.08$ ($p = 0.52 \pm 0.16$) in the Poppenhaeger et al. (2010) sample, finding a weak $\log(R'_{\rm HK}) - M_{\rm P}$ correlation ($p \approx 0.3$). They also found evidence that massive close-in planets are found around

stars with enhanced chromospheric activity but attributed this to selection effects. Krejčová & Budaj (2012) found stars hosting more massive planets show higher activity but noted the correlation may be affected by selection effects. Fossati et al. (2015a) found a $\log(R'_{\rm HK}) - M_{\rm P}$ correlation of $\rho = 0.40$, which marginally increased to $\rho = 0.44$ for $a < 0.06$ AU systems, but suggested these trends may have their origin in $\log(R'_{\rm HK}) - M_\star$ and $M_\star - M_{\rm P}$ correlations. Caution is thus required when drawing inferences from apparent $\log(R'_{\rm HK}) - M_{\rm P}$ correlations.

Activity–mass relationship in the OU-SALT sample

Figure 5.12(a) shows the $\log(R'_{\rm HK}) - \log M_{\rm P}$ distribution for the OU SALT sample, with an equilibrium temperature colour scale. There is a positive $\log(R'_{\rm HK}) - \log M_{\rm P}$ correlation in the high activity MS population ($\rho = 0.36 \pm 0.03$, $p = 0.02 \pm 0.012$), which strengthens when only the HJ population is fitted ($\rho = 0.52 \pm 0.01$, $p = 0.01 \pm 0.002$). This suggests higher mass planets orbit more active stars. The low activity correlation is weaker ($\rho = 0.29 \pm 0.02$, $p = 0.08 \pm 0.021$). The correlations are stronger than those reported by Martins et al. (2011) and approximately equivalent to the HJ-host population studied by Fossati et al. (2015b). A non-significant negative correlation is seen in the SG population ($\rho = 0.29 \pm 0.02$, $p = 0.08 \pm 0.021$). For the $M_{\rm P} > 0.8\ M_{\rm J}$ and $a < 0.05$ AU sub-samples, the correlations strengthen markedly to $\rho = 0.55 \pm 0.07$ ($p = 0.05 \pm 0.048$) and $\rho = 0.64 \pm 0.06$ ($p = 0.04 \pm 0.031$) respectively. Massive, close-in planets have strong gravitational and radiative interactions with their hosts, making SPI and mass-loss more likely.

5.5 SPI proxies

Proxy expressions may be used to approximate SPI strength. For example, Miller et al. (2015) use $M_{\rm P} \sin i / a^2$ as a magnetic SPI proxy. This does not however distinguish between magnetic and tidal SPI effects as $M_{\rm P}$ strongly correlates with $h_{\rm tide}/h_{\rm scale}$ (Staab et al., 2017) . Moreover, close-in planets may not be subject to magnetic fields that drop at a^{-2} as the geometry of close-in fields may change (Lanza, 2013).

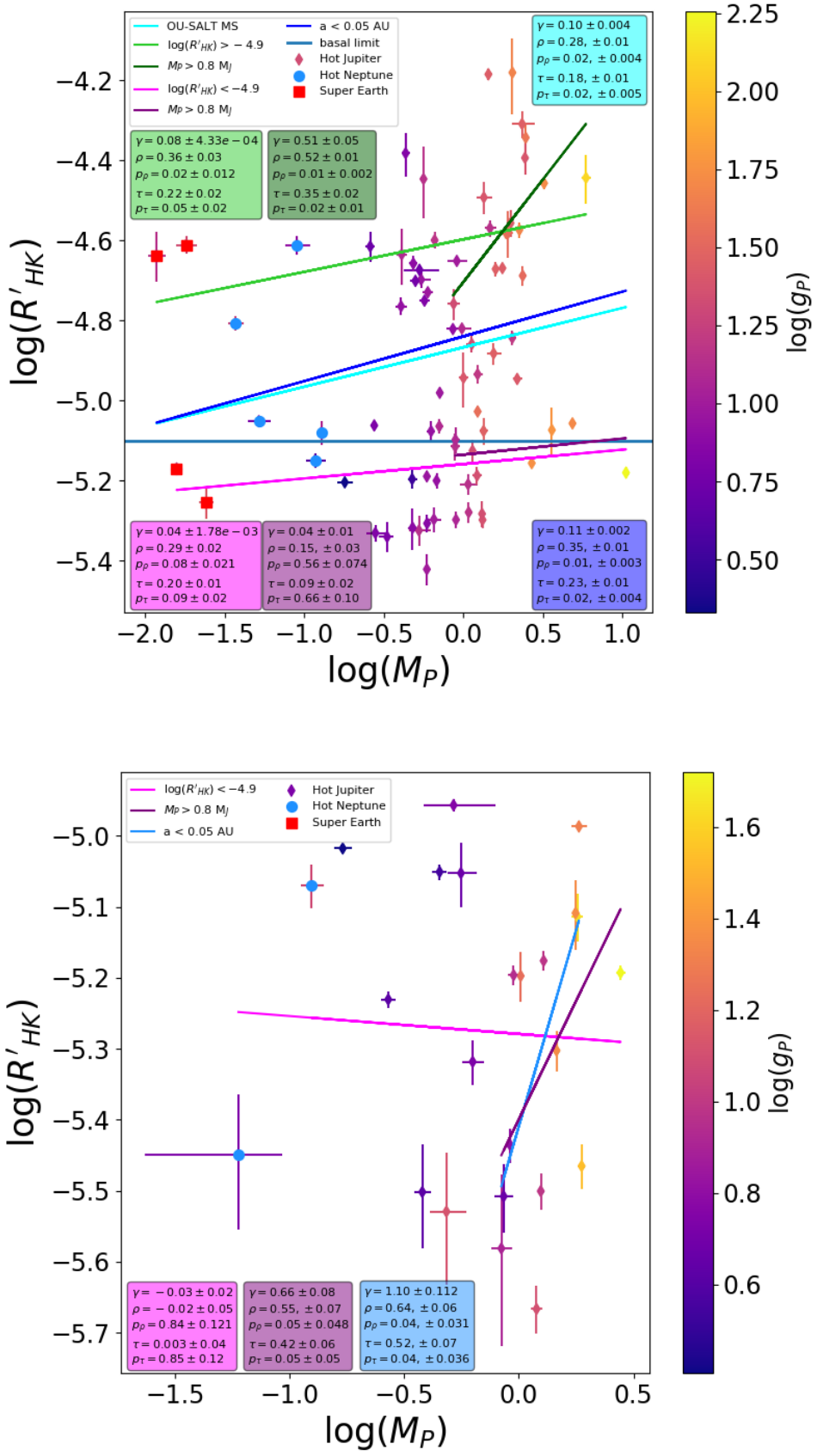

FIGURE 5.12: Activity–planetary mass distributions for MS (top panel) and SG (bottom panel) samples, with a colour scale showing the log of surface gravity. Linear fits for a variety of subsamples are plotted as detailed in the keys.

Studies have searched for dependencies between stellar activity and SPMI proxies M_P/a and M_P/a^2. In a sample of 72 stars, Poppenhaeger et al. (2010) identified a correlation between X-ray luminosity and M_P/a but ascribed the dependence to selection effects. They define "close-in heavy planets" with $\log(M_P/a) > 0.5$, which correspond to Jupiter-like bodies at a maximum orbital distance of ~ 0.3 AU. Miller et al. (2015) found correlations between X-ray luminosity measurements and M_P/a^2 but identified selection effects, and attributed the trend to a subset of "extreme" systems with $M_P/a^2 > 450$ M_J AU^2. Activity was consistent with rotation rate, so did not seem to be influenced by SPMI. France et al. (2018) identified a correlation between FUV activity indices and M_P/a for a sample of 71 planet hosts. However, once parameter interdependency was accounted for, no significant SPI signal remained.

Figure 5.13 shows the OU-SALT $\log(R'_{HK}) - \log(M_P/a)$ distribution, with linear fits either side of $\log(M_P/a) = 0.6$ for the MS sample, and $\log(M_P/a) = 1.2$ for the SG sample. There is a strong positive $\log(R'_{HK}) - \log(M_P/a)$ correlation in the MS where $\log(M_P/a) > 0.6$ ($\rho = 0.42 \pm 0.01$, $p = (5.3 \pm 1.4) \times 10^{-4}$). The less massive MS population (of super Earths & mini Neptunes) has a non-significant, negative correlation ($\rho = -0.53 \pm 0.03, p = 0.10 \pm 0.018$). A stronger correlation is present in the SG sample for systems with $\log(M_P/a) > 1.2$ ($\rho = 0.49 \pm 0.08$, $p = 0.086 \pm 0.061$). Activity appears strongly correlated to M_P/a for close-in, heavy planets.

Figure 5.14 shows M_P/a^2 plotted against activity. A strong positive $\log(R'_{HK}) - M_P/a^2$ correlation is present in the MS sample ($\rho = 0.34 \pm 0.01$, $p = 0.002 \pm 5.4 \times 10^{-4}$). The correlation is negative and gradient weak in the SG population, with $\rho = -0.08 \pm 0.03$ ($p = 0.73 \pm 0.11$). Correlations strengthen for the 'extreme' ($M_P/a^2 > 450$ M_J AU^{-2}) sub-samples in both MS and SG populations to $\rho = 0.45 \pm 0.01$ ($p = 0.006 \pm 0.002$) and 0.59 ± 0.06 ($p = 0.04 \pm 0.02$) respectively. Figures 5.13 and 5.14 thus show correlations between activity and SPMI proxies that strengthen for sub-samples that contain more massive systems on closer-in orbits. Although positive correlations between the SPI proxies and activity indicate the possible presence of observable SPI in the OU-SALT sample, interdependencies and selection effects must be accounted for before firm conclusions are drawn (Poppenhaeger et al., 2010; Miller et al., 2015; France et al., 2018).

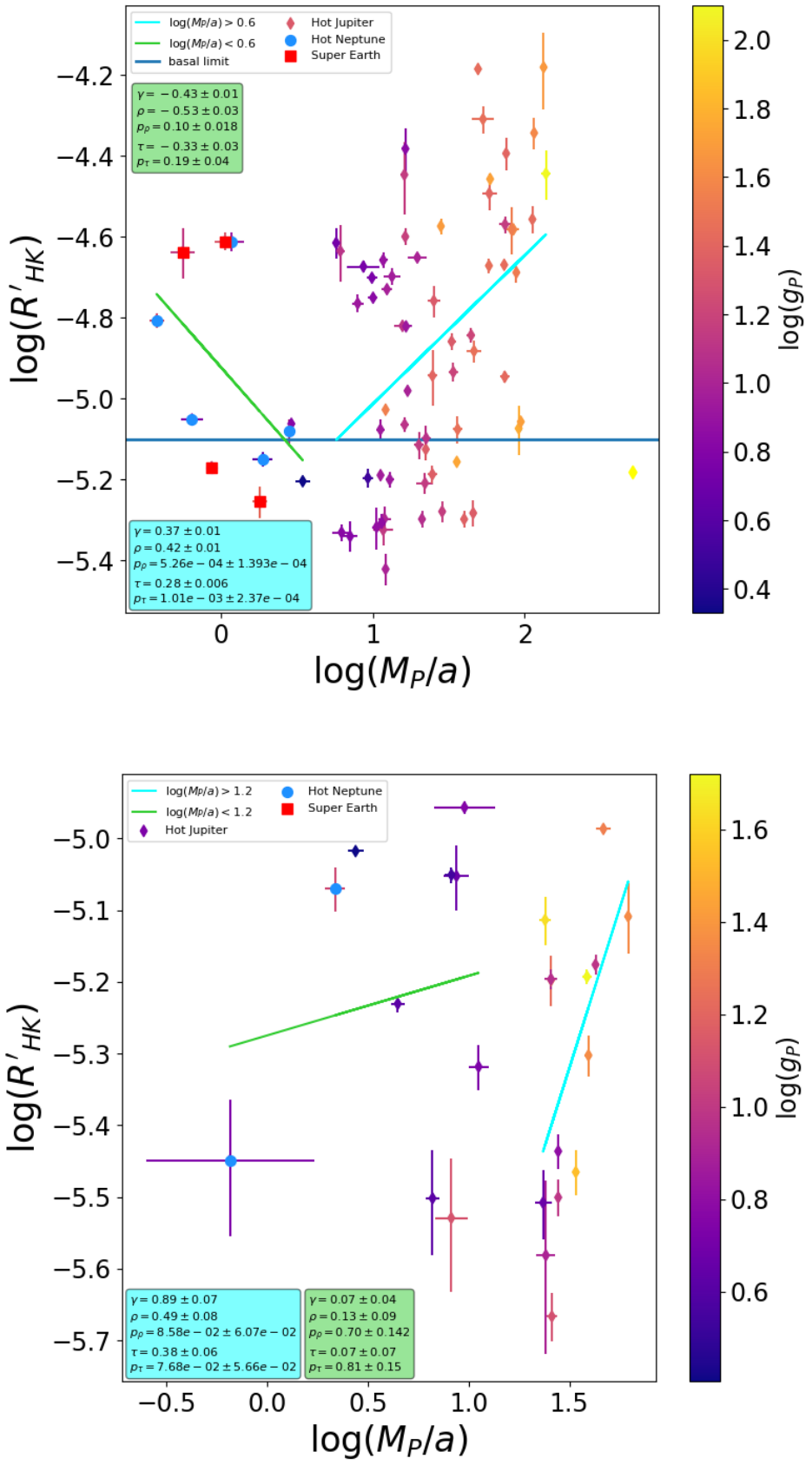

FIGURE 5.13: The logarithm of (M_P/a) plotted against activity for MS (top) and SG (bottom) samples. Each plot shows correlations for sub-samples as detailed.

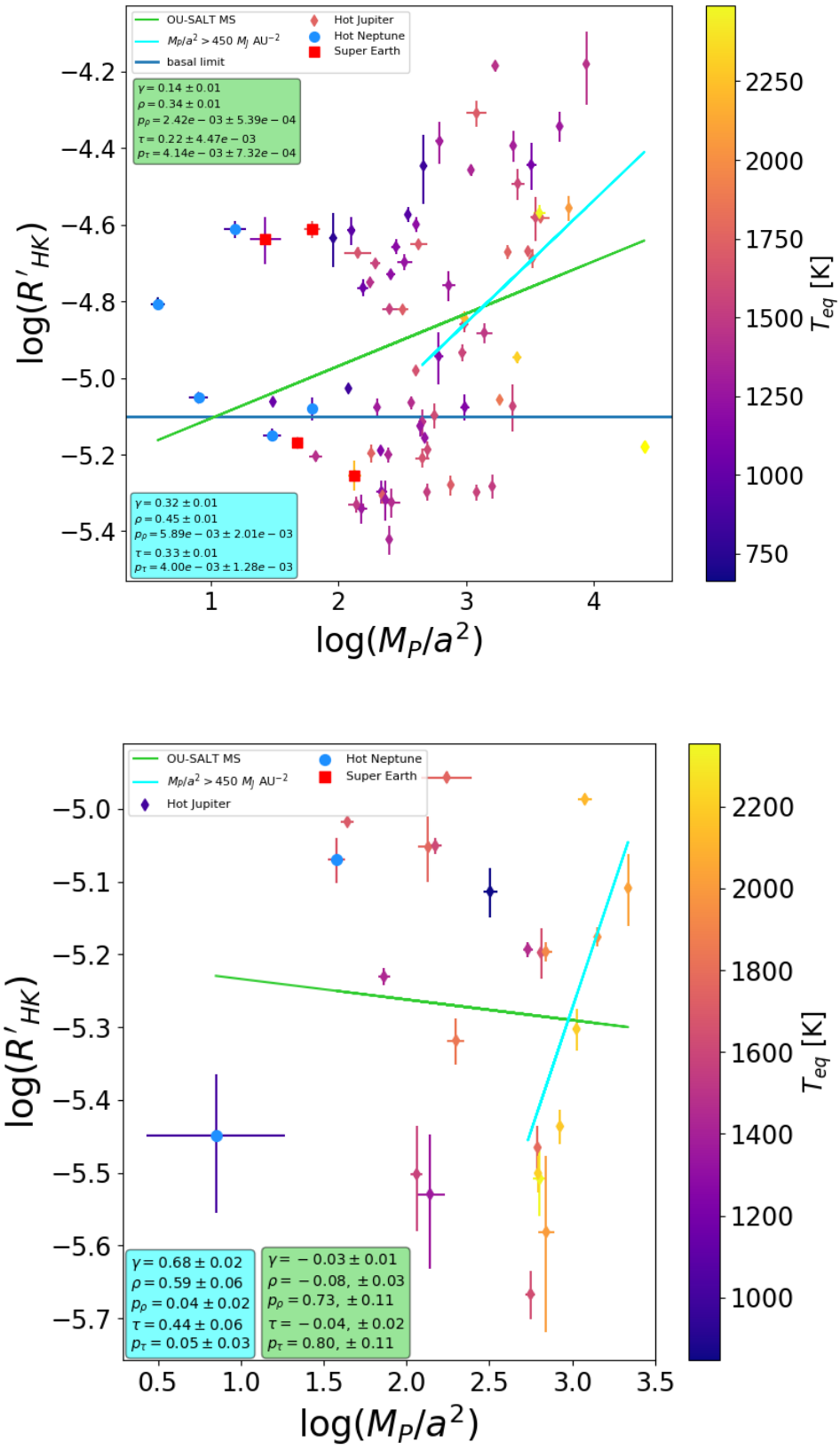

FIGURE 5.14: Magnetic SPI proxy $\log(M_P/a^2)$ plotted against activity. Each plot shows correlations for sub-populations as detailed.

TABLE 5.2: Planetary parameters and SPI terms examined in this Chapter: surface gravity (g_P), density(ρ), equilibrium temperature (T_{eq}), tidal bulge height as a fraction of scale height (h_{tide}/h_{scale}), angular momentum of stellar rotation ($J_\star$), and angular momentum of planetary orbit (J_{orb}).

System	g_P	ρ	T_{eq}	h_{tide}/h_{scale}	$J_\star$	J_{orb}
	[ms^{-2}]	[$\rho_\odot$]	[K]		[kg m^2 d^{-1}]	[kg m^2 d^{-1}]
CoRoT-7	$22^{+3.5}_{-3.0}$	$5.7^{+1.10}_{-1.10}$	1760±30	0.004 49	1.25×10^{47}	1.64×10^{45}
CoRoT-11	$29^{+4.2}_{-4.2}$	$0.80^{+0.12}_{-0.12}$	1710±53	0.058 40	3.31×10^{48}	4.04×10^{47}
CoRoT-22	$5.4^{+5.7}_{-5.7}$	$1.0^{+1.00}_{-1.00}$	1010±45	0.000 09	2.39×10^{47}	1.39×10^{46}
CoRoT-28	$13^{+5.0}_{-3.6}$	$0.50^{+0.20}_{-0.20}$	1360±59	0.010 10	2.58×10^{47}	8.64×10^{46}
EPIC 219388192	1000^{+120}_{-120}	$45^{+5.70}_{-5.70}$	1160±35			
HAT-P-27	$15^{+1.7}_{-1.7}$	$0.60^{+0.10}_{-0.10}$	1210±39	0.009 51	1.62×10^{46}	9.42×10^{46}
HATS-2	$25^{+3.0}_{-3.0}$	$0.80^{+0.11}_{-0.11}$	1580±35	0.139 00	5.69×10^{46}	1.40×10^{47}
HATS-3	$14^{+2.0}_{-2.0}$	$0.40^{+0.05}_{-0.05}$	1650±28	0.019 70	7.41×10^{47}	1.90×10^{47}
HATS-10	$14^{+2.9}_{-2.4}$	$0.50^{+0.11}_{-0.11}$	1410±47	0.007 70	3.31×10^{47}	8.56×10^{46}
HATS-13	$9.1^{+1.4}_{-1.2}$	$0.30^{+0.04}_{-0.04}$	1250±21	0.007 08	1.15×10^{47}	7.84×10^{46}
HATS-18	$27^{+2.3}_{-3.7}$	$0.80^{+0.10}_{-0.15}$	2060±74	0.767 00	3.15×10^{47}	1.96×10^{47}
HATS-21	$6.5^{+1.1}_{-1.4}$	$0.20^{+0.05}_{-0.07}$	1280±58	0.003 64	1.39×10^{47}	5.46×10^{46}
HATS-27	$5.6^{+2.3}_{-1.6}$	$0.10^{+0.06}_{-0.06}$	1660±83	0.007 43	1.10×10^{48}	1.14×10^{47}
HATS-29	$10^{+1.2}_{-1.1}$	$0.30^{+0.04}_{-0.04}$	1210±33	0.004 77	1.25×10^{47}	1.14×10^{47}
HATS-30	$13^{+1.4}_{-1.2}$	$0.40^{+0.06}_{-0.06}$	1420±32	0.010 30	2.28×10^{47}	1.13×10^{47}
HATS-36	$52^{+3.2}_{-3.0}$	$1.6^{+0.15}_{-0.15}$	1370±29	0.026 60	3.36×10^{47}	6.06×10^{47}
HD 73256			1320			
K2-2	$18^{+3.7}_{-3.1}$	$3.1^{+0.81}_{-0.81}$	663±16	0.000 02	5.31×10^{46}	1.04×10^{46}
K2-19	$8.4^{+1.2}_{-2.7}$	$0.30^{+0.06}_{-0.06}$	880±26	0.000 18	7.66×10^{46}	1.76×10^{46}
K2-31	$39^{+26.4}_{-26.4}$	$1.0^{+1.00}_{-1.00}$	1750±95	0.274 00	1.23×10^{47}	1.91×10^{47}
K2-32	$6.1^{+1.2}_{-1.2}$	$0.60^{+0.19}_{-0.19}$	825±24	0.000 07	2.42×10^{46}	9.96×10^{45}
K2-39	$12^{+3.0}_{-3.0}$	$0.90^{+0.35}_{-0.24}$	1690±66	0.011 60	1.67×10^{46}	2.39×10^{46}
K2-96	$17^{+3.6}_{-3.6}$	$4.2^{+1.60}_{-1.10}$	1750±47	0.003 23	6.02×10^{46}	1.45×10^{45}
K2-106	$27^{+3.6}_{-3.9}$	$9^{+3.60}_{-2.40}$	2230±63	0.022 80	1.17×10^{47}	2.04×10^{45}
K2-135	$14^{+3.9}_{-3.0}$	$3.6^{+1.50}_{-1.00}$	1120±91	0.000 87	3.96×10^{46}	1.15×10^{45}
KELT-10	$8.6^{+0.7}_{-0.8}$	$0.20^{+0.03}_{-0.03}$	1380±34	0.007 20	2.32×10^{47}	1.20×10^{47}
KELT-11	$2.6^{+0.5}_{-0.5}$	$0.10^{+0.02}_{-0.02}$	1700±19	0.007 62	4.93×10^{47}	3.74×10^{46}
KELT-15	$13^{+2.1}_{-1.8}$	$0.30^{+0.08}_{-0.06}$	1640±54	0.037 30	6.36×10^{47}	2.04×10^{47}
Qatar-2	$39^{+0.5}_{-0.5}$	$1.8^{+0.03}_{-0.03}$	1290±17	0.166 00	8.03×10^{46}	2.25×10^{47}
WASP-5	$29^{+2.6}_{-2.6}$	$0.90^{+0.12}_{-0.12}$	1750±40	0.131 00	1.64×10^{47}	1.97×10^{47}
WASP-6	$8.0^{+0.3}_{-0.3}$	$0.20^{+0.01}_{-0.01}$	1180±27	0.005 73	5.53×10^{46}	6.61×10^{46}
WASP-7	$13^{+2.4}_{-2.4}$	$0.40^{+0.09}_{-0.09}$	1530±50	0.008 52	1.58×10^{48}	2.06×10^{47}
WASP-8	$49^{+2.1}_{-2.1}$	$1.0^{+1.00}_{-1.00}$	928±30	0.003 68	7.41×10^{46}	4.73×10^{47}

System	g_P [ms^{-2}]	ρ [$\rho_\odot$]	T_{eq} [K]	h_{tide}/h_{scale}	$J_\star$ [kg m^2 d^{-1}]	J_{orb} [kg m^2 d^{-1}]
WASP-15	$7.4^{+0.5}_{-0.5}$	$0.20^{+0.02}_{-0.02}$	1720±33	0.010 50	3.80×10^{47}	1.12×10^{47}
WASP-17	$3.2^{+0.2}_{-0.2}$	$0.10^{+0.00}_{-0.00}$	1750±39	0.009 55	8.77×10^{47}	9.00×10^{46}
WASP-18	$180^{+6.4}_{-6.4}$	$5.6^{+0.31}_{-0.31}$	2410±42	3.610 00	8.56×10^{47}	1.26×10^{48}
WASP-25	$9.5^{+0.8}_{-0.8}$	$0.30^{+0.03}_{-0.03}$	1210±16	0.004 69	1.40×10^{47}	9.85×10^{46}
WASP-26	$17^{+1.3}_{-1.3}$	$0.50^{+0.06}_{-0.06}$	1650±30	0.032 50	2.62×10^{47}	1.55×10^{47}
WASP-28	$14^{+0.9}_{-0.9}$	$0.50^{+0.03}_{-0.03}$	1450±26	0.012 80	1.67×10^{47}	1.36×10^{47}
WASP-32	$59^{+5.4}_{-5.4}$	$2.2^{+0.40}_{-0.40}$	1560±44	0.080 80	2.81×10^{47}	5.49×10^{47}
WASP-34	$9.1^{+1.1}_{-1.4}$	$0.30^{+0.08}_{-0.08}$	1160±75	0.003 54	6.29×10^{46}	1.00×10^{47}
WASP-36	$33^{+1.1}_{-1.1}$	$0.90^{+0.04}_{-0.04}$	1730±23	0.158 00	1.68×10^{47}	2.94×10^{47}
WASP-38	$56^{+3.2}_{-3.2}$	$2.1^{+0.14}_{-0.14}$	1250±22	0.010 00	6.59×10^{47}	5.92×10^{47}
WASP-43	$48^{+1.4}_{-1.3}$	$1.8^{+0.08}_{-0.08}$	1440±41	0.495 00	9.16×10^{46}	1.56×10^{47}
WASP-44	$22^{+1.6}_{-1.6}$	$0.80^{+0.07}_{-0.07}$	1310±51	0.019 00	1.21×10^{47}	1.14×10^{47}
WASP-45	$25^{+1.3}_{-1.3}$	$1.0^{+0.08}_{-0.08}$	1170±50	0.015 50	9.13×10^{46}	1.40×10^{47}
WASP-46	$34^{+1.1}_{-1.1}$	$1.1^{+0.05}_{-0.05}$	1640±57	0.154 00	6.46×10^{46}	1.95×10^{47}
WASP-47	$22^{+0.3}_{-0.3}$	$0.70^{+0.02}_{-0.02}$	1270±18	0.012 20	1.02×10^{47}	1.92×10^{47}
WASP-50	$28^{+0.6}_{-0.6}$	$0.90^{+0.03}_{-0.03}$	1440±25	0.053 20	9.16×10^{46}	1.67×10^{47}
WASP-51	$9.8^{+0.9}_{-0.9}$	$0.30^{+0.04}_{-0.04}$	1650±38	0.015 40	1.59×10^{47}	1.19×10^{47}
WASP-52	$6.9^{+0.3}_{-0.3}$	$0.20^{+0.01}_{-0.01}$	1320±32	0.020 20	1.09×10^{47}	4.63×10^{46}
WASP-55	$8.7^{+0.5}_{-0.4}$	$0.20^{+0.02}_{-0.02}$	1300±18	0.004 29	1.90×10^{47}	1.17×10^{47}
WASP-62	$6.8^{+0.7}_{-0.6}$	$0.20^{+0.03}_{-0.03}$	1430±34	0.005 05	6.66×10^{47}	1.11×10^{47}
WASP-63	$4.2^{+0.5}_{-0.5}$	$0.10^{+0.02}_{-0.02}$	1580±46	0.009 06	3.33×10^{47}	7.64×10^{46}
WASP-65	$29^{+0.9}_{-1.2}$	$1.1^{+0.07}_{-0.08}$	1490±59	0.053 30	1.62×10^{47}	2.00×10^{47}
WASP-67	$8.5^{+0.8}_{-0.8}$	$0.30^{+0.04}_{-0.04}$	1050±25	0.002 12	6.81×10^{46}	6.11×10^{46}
WASP-69	$5.3^{+0.6}_{-0.6}$	$0.20^{+0.03}_{-0.03}$	961±20	0.002 28	7.07×10^{46}	3.68×10^{46}
WASP-70	$10^{+1.6}_{-1.1}$	$0.40^{+0.10}_{-0.06}$	1380±55	0.008 64	1.16×10^{47}	1.00×10^{47}
WASP-72	$23^{+7.3}_{-5.5}$	$0.70^{+0.43}_{-0.25}$	2200±139	0.190 00	7.88×10^{47}	2.42×10^{47}
WASP-73	$35^{+5.1}_{-5.8}$	$1.2^{+0.26}_{-0.30}$	1780±90	0.062 60	8.10×10^{47}	3.73×10^{47}
WASP-74	$8.9^{+0.4}_{-0.4}$	$0.20^{+0.02}_{-0.02}$	1920±53	0.075 10	4.76×10^{47}	1.63×10^{47}
WASP-75	$15^{+0.3}_{-0.3}$	$0.50^{+0.02}_{-0.02}$	1710±34	0.039 40	3.03×10^{47}	1.65×10^{47}
WASP-76	$6.3^{+0.3}_{-0.3}$	$0.20^{+0.01}_{-0.01}$	2180±46	0.127 00	3.99×10^{47}	1.48×10^{47}
WASP-77	$28^{+0.5}_{-0.5}$	$1.0^{+0.03}_{-0.03}$	1710±22	0.121 00	1.83×10^{47}	1.99×10^{47}
WASP-78	$4.7^{+0.5}_{-0.5}$	$0.10^{+0.01}_{-0.01}$	2350±81	0.203 00	1.13×10^{48}	1.42×10^{47}
WASP-79	$8.4^{+0.7}_{-0.7}$	$0.20^{+0.02}_{-0.02}$	1720±37	0.014 70	1.92×10^{48}	1.69×10^{47}
WASP-80	$14^{+0.5}_{-0.5}$	$0.50^{+0.02}_{-0.02}$	825±23	0.006 46	2.15×10^{46}	5.92×10^{46}
WASP-82	$9.8^{+0.7}_{-0.8}$	$0.20^{+0.03}_{-0.03}$	2210±58	0.103 00	4.52×10^{47}	2.47×10^{47}
WASP-87 A	$26^{+1.9}_{-1.9}$	$0.80^{+0.09}_{-0.09}$	2310±68	0.010 20	1.21×10^{48}	1.25×10^{47}
WASP-88	$4.7^{+0.8}_{-0.8}$	$0.10^{+0.03}_{-0.03}$	1760±63	0.379 00	9.00×10^{47}	3.00×10^{47}

System	g_P [ms^{-2}]	ρ [$\rho_\odot$]	T_{eq} [K]	h_{tide}/h_{scale}	$J_\star$ [kg m^2 d^{-1}]	J_{orb} [kg m^2 d^{-1}]
WASP-89	$130^{+5.8}_{-5.8}$	$5.3^{+0.33}_{-0.33}$	1090±32	0.071 60	9.69×10^{46}	8.56×10^{47}
WASP-90	$5.4^{+0.8}_{-0.7}$	$0.10^{+0.03}_{-0.03}$	1840±59	0.016 40	8.81×10^{47}	1.36×10^{47}
WASP-91	$29^{+2.1}_{-1.9}$	$1.2^{+0.10}_{-0.10}$	1140±31	0.024 70	8.30×10^{46}	1.75×10^{47}
WASP-94A	$3.5^{+0.2}_{-0.2}$	$0.10^{+0.01}_{-0.01}$	1620±36	0.008 00	4.72×10^{47}	9.25×10^{46}
WASP-94 B			1850±88			
WASP-95	$22^{+1.0}_{-3.3}$	$0.80^{+0.07}_{-0.02}$	1620±72	0.045 90	1.86×10^{47}	1.61×10^{47}
WASP-96	$7.6^{+0.7}_{-0.7}$	$0.30^{+0.04}_{-0.04}$	1280±50	0.006 47	7.99×10^{46}	7.68×10^{46}
WASP-97	$23^{+2.2}_{-2.2}$	$0.90^{+0.11}_{-0.11}$	1550±44	0.051 60	6.25×10^{46}	1.86×10^{47}
WASP-99	$53^{+3.8}_{-6.8}$	$2.1^{+0.30}_{-0.30}$	1470±54	0.024 40	8.48×10^{47}	6.64×10^{47}
WASP-100	$16^{+6.7}_{-4.7}$	$0.40^{+0.20}_{-0.20}$	2200±171	0.111 00	1.92×10^{48}	3.98×10^{47}
WASP-101	$5.8^{+0.6}_{-0.5}$	$0.20^{+0.02}_{-0.02}$	1550±40	0.006 55	1.03×10^{48}	9.54×10^{46}
WASP-103	$14^{+0.8}_{-0.8}$	$0.30^{+0.03}_{-0.03}$	2490±88	0.926 00	8.64×10^{47}	1.66×10^{47}
WASP-104	$23^{+1.2}_{-1.2}$	$1.0^{+0.03}_{-0.03}$	1510±40	0.067 40	1.82×10^{46}	1.63×10^{47}
WASP-105	$44^{+1.0}_{-1.0}$	$2.0^{+0.10}_{-0.10}$	847±31	0.003 61	6.52×10^{46}	3.44×10^{47}
WASP-108	$12^{+0.7}_{-0.7}$	$0.40^{+0.03}_{-0.03}$	1600±50	0.024 70	3.26×10^{47}	1.40×10^{47}
WASP-109	$10^{+1.5}_{-1.5}$	$0.30^{+0.05}_{-0.05}$	1700±50	0.017 50	1.19×10^{48}	1.57×10^{47}
WASP-111	$20^{+2.5}_{-2.5}$	$0.60^{+0.11}_{-0.11}$	2120±80	0.153 00	1.49×10^{48}	3.24×10^{47}
WASP-114	$23^{+1.9}_{-1.8}$	$0.70^{+0.10}_{-0.10}$	2030±66	0.299 00	5.65×10^{47}	2.48×10^{47}
WASP-117	$6.6^{+0.9}_{-0.8}$	$0.30^{+0.05}_{-0.05}$	1020±33	0.000 33	1.05×10^{47}	6.57×10^{46}
WASP-118	$5.7^{+0.6}_{-0.5}$	$0.20^{+0.07}_{-0.07}$	1750±36	0.011 90	1.07×10^{48}	1.02×10^{47}
WASP-119	$16^{+4.1}_{-3.3}$	$0.50^{+0.20}_{-0.20}$	1570±72	0.045 10	4.10×10^{46}	1.73×10^{47}
WASP-120	$51^{+5.6}_{-5.0}$	$1.5^{+0.33}_{-0.26}$	1880±66	0.149 00	1.88×10^{48}	9.48×10^{47}
WASP-122	$9.7^{+0.5}_{-0.5}$	$0.20^{+0.02}_{-0.02}$	1960±50	0.189 00	2.97×10^{47}	1.81×10^{47}
WASP-123	$12^{+1.1}_{-1.0}$	$0.40^{+0.06}_{-0.06}$	1520±48	0.023 30	7.17×10^{46}	1.47×10^{47}
WASP-126	$6.8^{+1.6}_{-1.3}$	$0.30^{+0.08}_{-0.08}$	1490±65	0.005 84	3.40×10^{46}	4.59×10^{46}
WASP-127	$2.1^{+0.3}_{-0.3}$	$0.10^{+0.01}_{-0.01}$	1430±30	0.002 81	2.16×10^{46}	3.11×10^{46}
WASP-130	$36^{+2.5}_{-2.4}$	$1.8^{+0.18}_{-0.18}$	832±21	0.000 99	2.39×10^{46}	2.91×10^{47}
WASP-131	$4.2^{+0.4}_{-0.4}$	$0.10^{+0.02}_{-0.02}$	1440±35	0.003 03	2.33×10^{47}	4.99×10^{46}
WASP-132	$13^{+1.2}_{-1.1}$	$0.60^{+0.06}_{-0.06}$	761±20	0.000 83	2.55×10^{46}	6.89×10^{46}
WASP-138	$23^{+2.2}_{-2.0}$	$0.90^{+0.10}_{-0.15}$	1590±40	0.019 60	6.12×10^{47}	2.19×10^{47}
WASP-139	$4.2^{+0.6}_{-0.5}$	$0.20^{+0.04}_{-0.04}$	918±32	0.000 31	1.48×10^{47}	2.03×10^{46}
WASP-140	$25^{+15.0}_{-9.0}$	$0.80^{+0.40}_{-0.40}$	1330±41	0.068 20	1.16×10^{47}	3.04×10^{47}
WASP-142	$8.1^{+1.2}_{-1.0}$	$0.20^{+0.05}_{-0.05}$	2010±73	0.083 10	3.24×10^{47}	1.32×10^{47}
WASP-156	$12^{+1.3}_{-1.3}$	$1.0^{+0.10}_{-0.10}$	970±25	0.000 92	1.16×10^{47}	1.83×10^{46}

5.6 Stellar parameters

Intrinsic chromospheric activity is directly related to the physical properties of stars. In this section, I will quantify the relationship between stellar parameters and activity in the OU-SALT population. It is important to consider whether correlations observed between activity and planetary parameters have their basis in the relationship between stellar and planetary variables. Tables 4.1 and 4.2 list stellar variable values used in this section.

5.6.1 Effective temperature

Effective temperature ($T_{\rm eff}$) dictates the level of convection in stars, and is highly correlated with activity. Figure 5.15 shows the $\log(R'_{\rm HK}) - T_{\rm eff}$ distribution for OU-SALT MS and SG samples. The expected relationship is evident in the MS sample ($\rho = -0.34 \pm 0.02$, $p = 3.16 \times 10^{-3} \pm 2.29 \times 10^{-3}$). The correlation is present despite any effects of mass-loss and SPI. It is an order of magnitude weaker in the SG population ($\rho = 0.032 \pm 0.044$, $p = 0.83 \pm 0.09$). While this may be due to a limited sample size, it further justifies our careful division of the OU-SALT population into MS and SG sub-samples. The $T_{\rm eq}$ colour scale show that hot stars host hot planets. Thus, the $\log(R'_{\rm HK}) - T_{\rm eff}$ relationship must be accounted for when examining $\log(R'_{\rm HK}) - T_{\rm eq}$ correlations.

5.6.2 Mass

The effective temperature of a star is closely related to its mass ($M_\star$: Hertzsprung, 1923; Russell, 1923; Popper, 1967, 1980; Ibanoglu et al., 2006; Eker et al., 2015). Activity and mass are correlated through the relationship between activity and effective temperature. The colour scale in Figure 5.16 indicates that more massive stars are hotter. In contrast to large surveys (such as Gray et al., 2006; Martins et al., 2011; Pace, 2013; Miller et al., 2015), a negative $\log(R'_{\rm HK}) - M_\star$ correlation ($\rho = -0.37 \pm 0.02$, $p = 1.29 \times 10^{-3} \pm 1.32 \times 10^{-3}$) is identified in the OU-SALT MS sample (Figure 5.16). Fossati et al. (2015a) identify a similar trend ($\rho = -0.51$) in a

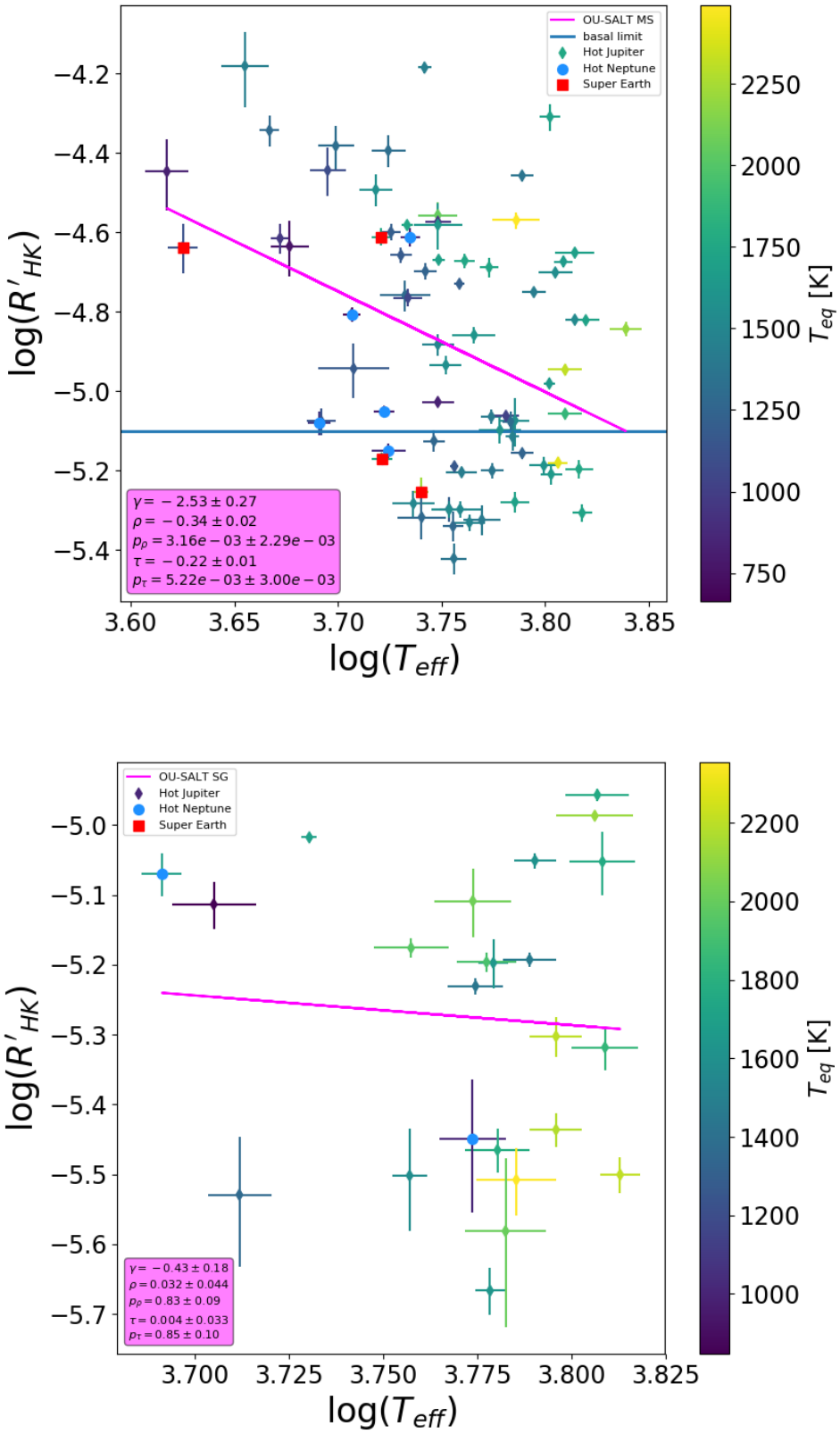

FIGURE 5.15: Effective temperature-activity distribution with equilibrium temperature colour scale for the MS (top panel) and SG (bottom panel) sub-samples.

sample of close-in transiting planet hosts. They note that high mass stars on average radiate more EUV flux, leading to higher rates of evaporation. As condensations of material escaped from transiting planets lie in the orbital plane, more absorption occurs around hotter stars. There is a weak positive $\log(R'_{\rm HK}) - M_\star$ correlation in the SG sample ($\rho = 0.032 \pm 0.044$) with a large p value (0.83 ± 0.09).

5.6.3 Rotation period

The relationship between chromospheric activity and stellar rotation period is discussed in detail in Section 1.3.4. Rotation periods are difficult to determine photometrically for low activity targets. Most OU-SALT hosts do not have published rotation periods. Where this is the case, I have estimated the maximum rotation period using the following expression:

$$\frac{P_{\rm rot}}{\sin i} = \frac{2\pi R_\star}{v_\star \sin i}, \tag{5.3}$$

where $R_\star$ is the stellar radius, $v_\star \sin i$ is the minimum rotational velocity of the star at the equator, and i is the (unknown) inclination.

Figure 5.17 shows $P_{\rm rot}/\sin i$ plotted against activity for MS and SG populations. A negative $\log(R'_{\rm HK}) - P_{\rm rot}/\sin i$ correlation is present in the MS population ($\rho = -0.19$, $p = 0.09$). This p-value is higher than for the stellar variable correlations above, which reflects the uncertainty of $P_{\rm rot}/\sin i$ estimations. Excluding two outliers, the trend is present and more significant in the SG population ($\rho = -0.31$, $p = 0.16$). There are a number of systems with relatively fast rotation but low activity. In particular, there are 11 'very fast rotators' ($P_{\rm rot} < 15$ d: McQuillan et al., 2014) with sub-basal activity. This likely results from enshrouding by material stripped from the star, and provides evidence that sub-basal systems are not just old stars. There are also a number of slower rotators (P > 30 days) with high activity levels ($\log(R'_{\rm HK})$ > -4.6). These hosts could be undergoing a degree of differential rotation that increases activity (Brandenburg & Giampapa, 2018).

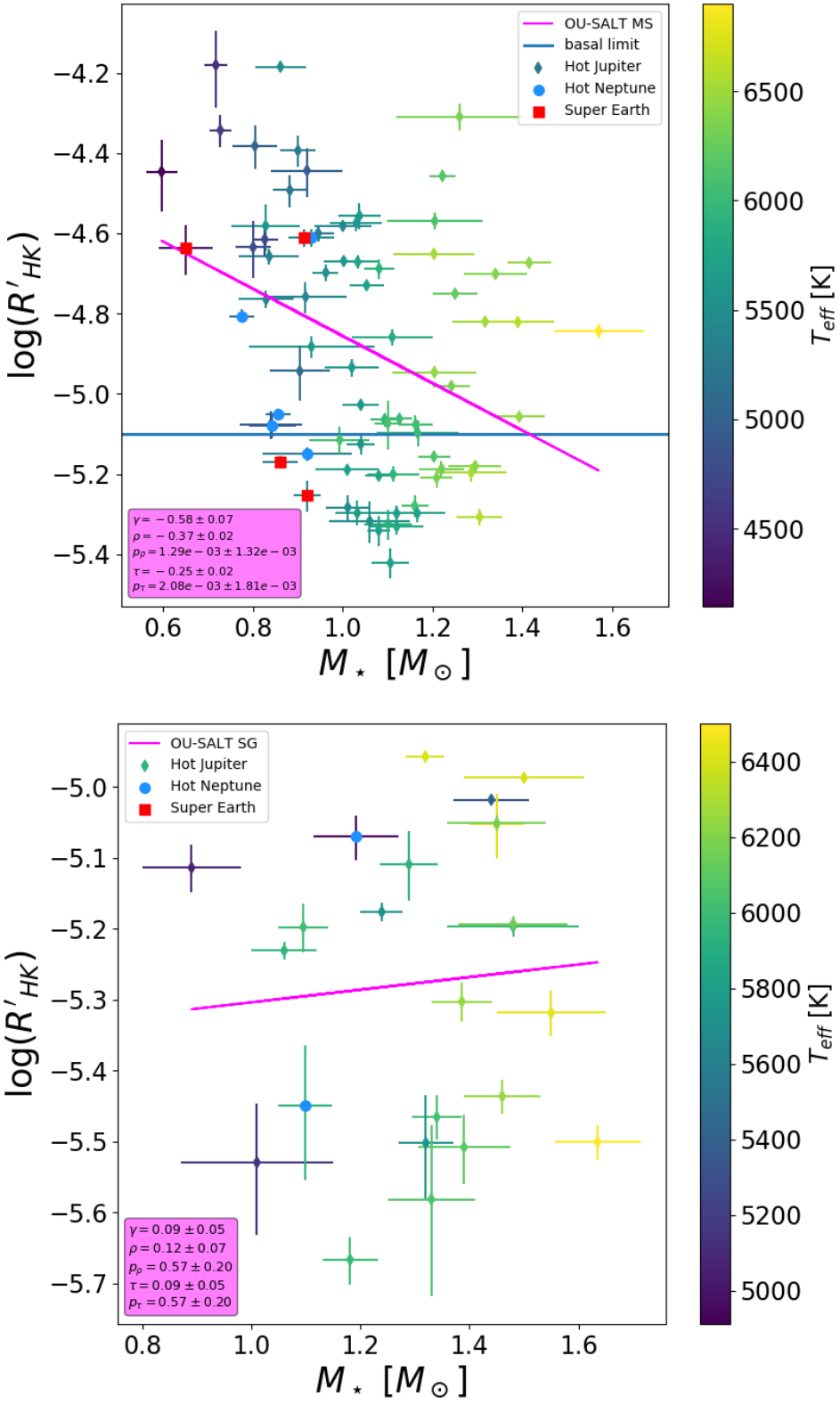

FIGURE 5.16: Stellar mass-activity relations with $T_{\rm eq}$ colour map for the MS (top panel) and SG (bottom panel) sub-samples.

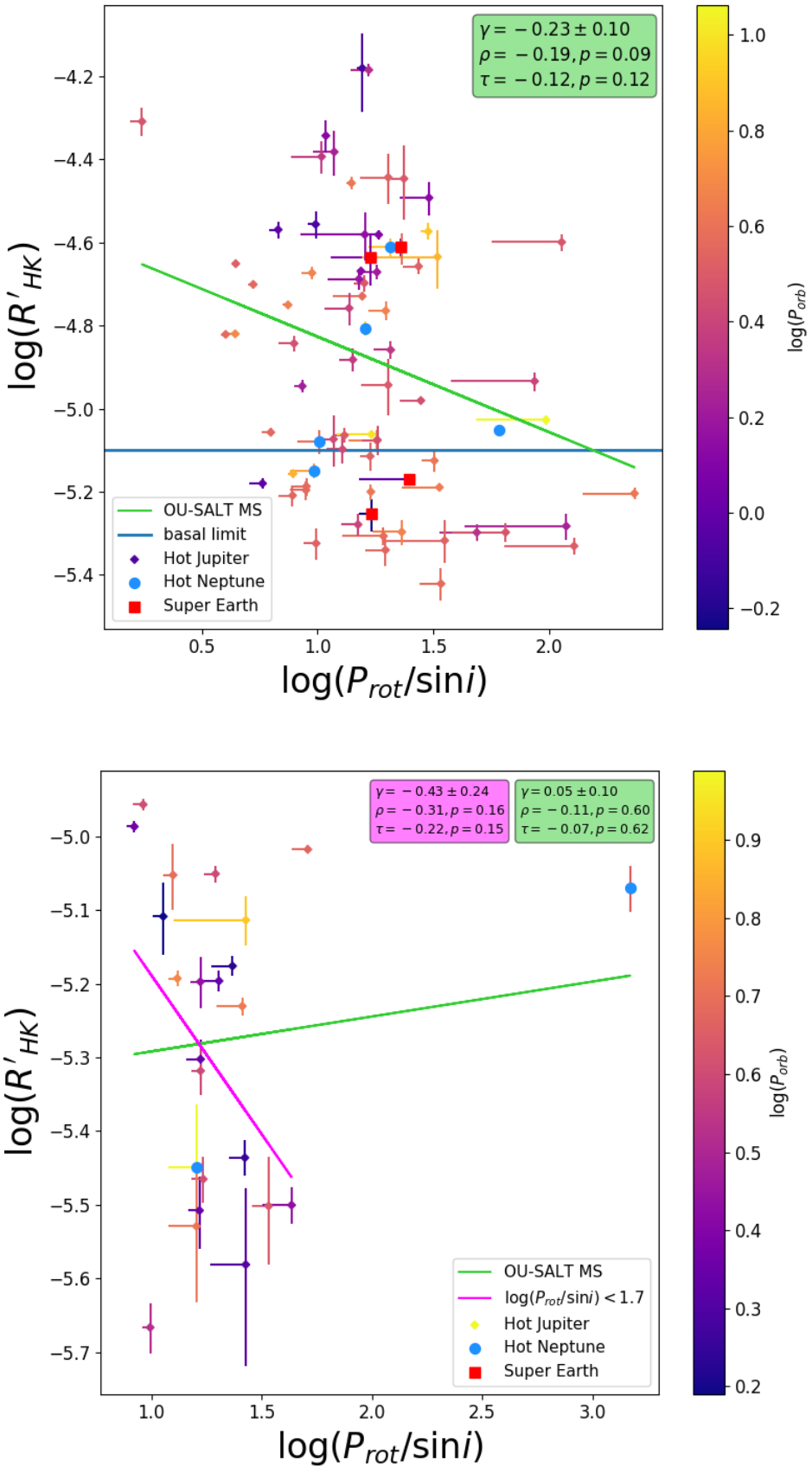

FIGURE 5.17: $P_{\rm rot}/\sin i$—being the upper limit for the rotation period of a star—plotted against activity for the MS (top panel) and SG (bottom panel) samples. The pink line in the lower panel excludes two outlying points with the longest rotation periods. The colour scale shows the logarithm of orbital period.

5.7 Interdependencies between variables

The relationships between activity, star mass, planet mass and semi-major axis are further explored in Figures 5.12, 5.18 and 5.19. Figure 5.12 shows a positive $\log(R'_{\rm HK}) - M_{\rm P}$ correlation ($\rho = 0.28 \pm 0.01$, $p = 0.02 \pm 0.004$), which remains relatively constant for systems with $a < 0.05$ AU ($\rho = 0.33 \pm 0.01$, $p = 0.031$). In the high-activity sample, this correlation strengthens for $M_{\rm P} > 0.8\ M_{\rm J}$ ($\rho = 0.52 \pm 0.01$, $p = 0.01 \pm 0.002$). There is a strong anti-correlation between the stellar/planet mass ratio and activity ($\rho = -0.34 \pm 0.01$, $p = 2.36 \times 10^{-3} \pm 5.47 \times 10^{-4}$) in Figure 5.18. A positive correlation between $\log M_\star$ and $\log M_{\rm P}$ ($\rho = 0.19 \pm 0.03$, $p = 0.19 \pm 0.06$) is present in Figure 5.19. Fossati et al. (2015b) identify a stronger activity–stellar/planet mass ratio anti-correlation ($\rho = -0.51$) in their population of close-in planet hosts. They argue the $\log(R'_{\rm HK}) - M_\star/M_{\rm P}$ correlation may be a consequence of the strong $\log(R'_{\rm HK}) - M_\star$ correlation. This may also be the case in the OU-SALT sample.

Figure 5.19 shows a strong $M_\star - a$ correlation ($\rho = 0.39 \pm 0.12$, $p = 3.81 \times 10^{-4} \pm 1.83 \times 10^{-4}$) that strengthens for $a < 0.05$ AU ($\rho = 0.49 \pm 0.01$, $p = 2.65 \times 10^{-4} \pm 2.43 \times 10^{-4}$). Fossati et al. (2015b) note a similar correlation in their sample of ~ 40 close-in planet hosts ($\rho = 0.33$) and a slightly stronger correlation for a sub-sample with $a < 0.06$ AU ($\rho = 0.48$), which they attribute to ground-based detection biases of transiting planets.

There are several alternative explanations. Massive planets may not survive close to massive stars, either disintegrating or being expelled from their close-in orbits. The planet formation process might also account for the correlation. Strong stellar magnetic field lines modify the accretion flow in the inner disk of young, low-mass stars evolving down their Hayashi tracks (Lépine & Gregorio-Hetem, 2003). The disk becomes truncated within a few stellar radii of the stellar surface (Camenzind, 1990; Koenigl, 1991). Disk material is channelled along the inner disk along magnetic field lines, and is no longer available for planet formation. The area of inner disk truncation increases with stellar mass. Hence, the closest-in planets are expected to have larger semi-major axes around more massive stars. Accretion funnel flows onto the star can cause accretion shocks that produce an inclined magnetosphere.

Inner disk warps induced by interaction with an inclined magnetosphere may halt the migration of planets to inner regions (Bouvier et al., 2006).

The $\log(R'_{\rm HK}) - a$ correlation may then find its origin in $M_\star - a$ and $\log(R'_{\rm HK}) - M_\star$ correlations. Based on this inference, Fossati et al. (2015b) conclude that the $\log(R'_{\rm HK}) - g_{\rm P}^{-1}$ correlation they identify is not caused by SPI but solely by planetary evaporation explanation—a conclusion that supports the findings of Poppenhaeger et al. (2010), Poppenhaeger & Schmitt (2011) and Miller et al. (2015). This highlights the importance of accounting for variation in stellar parameters in activity studies.

It is also important to note the relationship between stellar rotation period and planetary mass loss rate. Rotation rate is closely associated with activity and consequently extreme ultraviolet flux emission. This, in turn, correlates with mass loss rate (Kubyshkina et al., 2018a,b).

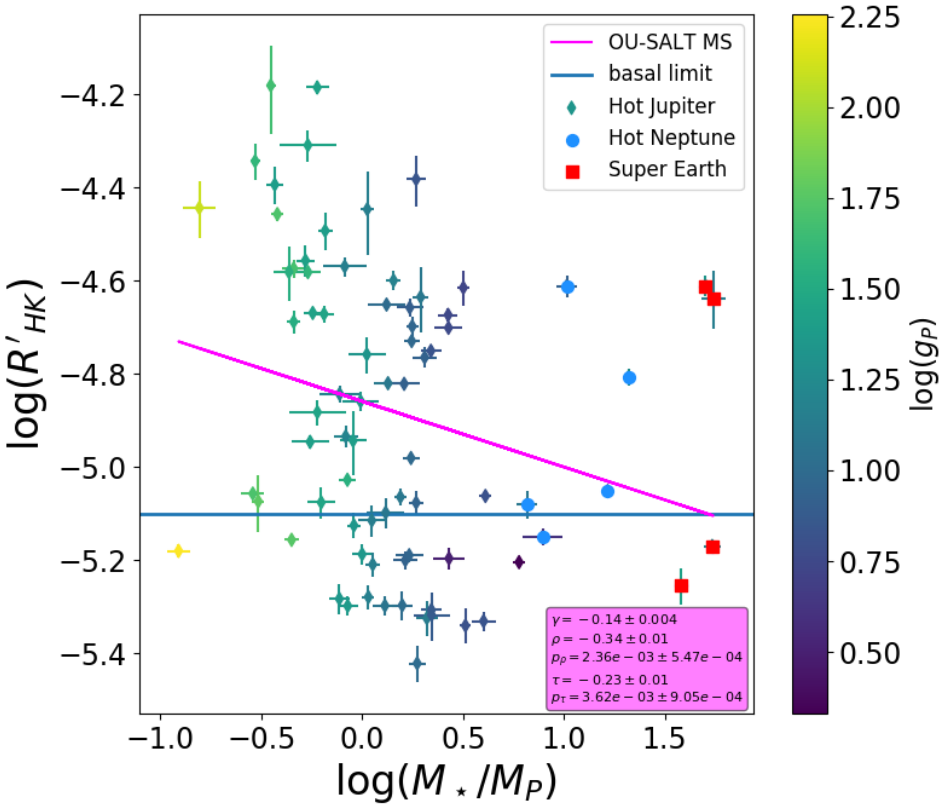

FIGURE 5.18: The $\log(R'_{\rm HK}) - \log(M_\star/M_{\rm P})$ distributions for OU-SALT MS sample.

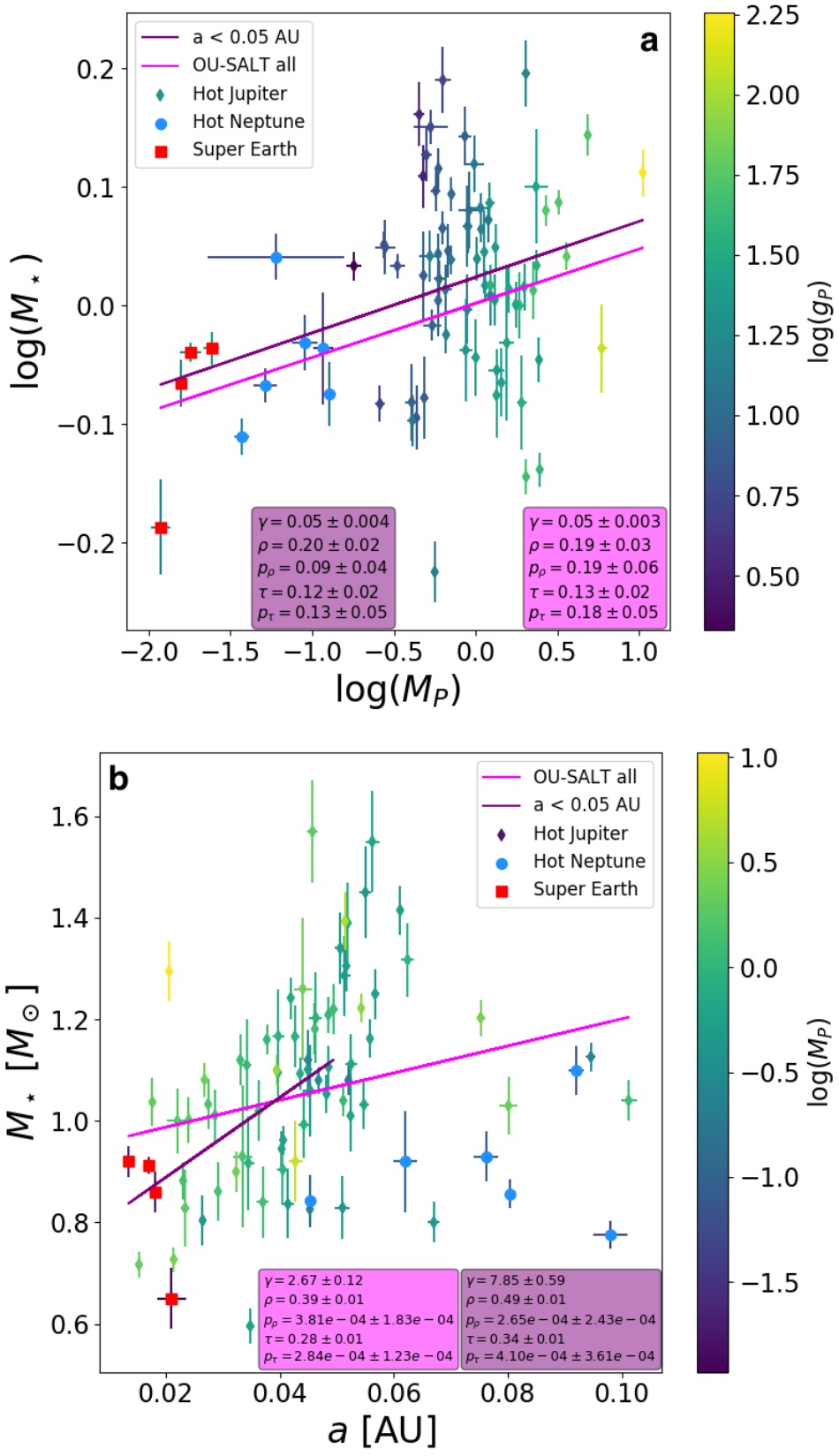

FIGURE 5.19: (a) $\log(M_\star) - \log(M_P)$ distribution ad (b) $M_\star - a$ distribution for the entire OU-SALT sample. Both plots have best fits lines for the full sample and an $a < 0.05$ AU sub-sample.

5.8 Constraining enshrouded systems

There are several competing factors that determine whether an ablated gas cloud or disk remains in a stable orbit around a host star. Stellar gravity is the main force by which a cloud remains bound to the star. On the other hand, the stellar wind—strongest in hot stars—may blow the disk away. To a lesser extent, radiation pressure may act to dissipate a circumstellar cloud (Carroll-Nellenback et al., 2017; Debrecht et al., 2018). In addition, the presence of a planetary magnetic field may disrupt the formation of a disk (Daley-Yates & Stevens, 2018).

To constrain the systems in which enshrouding occurs, I define a sample of 'absorbing systems' (Figure 5.20). This consists of hosts with sub-basal activity and, in addition, ten systems that drive the $\log(R'_{\rm HK}) - g_{\rm P}$ correlation in the high activity MS sample shown in Figure 5.8: namely HATS-13, HAT-27, K2-19, WASP-6, WASP-25, WASP-62, WASP-67, WASP-79, WASP-101 and WASP-109. When these 10 systems are removed from the high activity sub-sample, the $\log(R'_{\rm HK}) - g_{\rm P}$ slope decreases by 0.49 and ρ by 0.10 (Figure 5.20).

Figure 5.21 shows the $\log(R'_{\rm HK}) - (B-V)$ and $\log(R'_{\rm HK}) - M_\star$ distributions for the OU-SALT MS sample. The distribution of the absorbing and non-absorbing subsamples appears distinct in the lower panels. An Anderson Darling test score of 4.09 indicates there is a low probability ($p = 7.5 \times 10^{-3}$) that the non-absorbing and absorbing populations have been drawn from the same distribution. Absorption occurs only in systems where $B-V \lesssim 0.8$ (which equates to $T_{\rm eff} \gtrsim 5200$ K). At lower temperatures, EUV flux may be insufficient to cause a blow-off, or thermal energy may be inadequate to support a dispersed disk. EUV flux becomes highly efficient at stripping mass in hotter systems. There is an upper temperature boundary for disk stability at $B-V \sim 0.4$ ($T_{\rm eff} \sim 6600$ K). At these temperatures, the solar wind may become sufficiently strong to destabilise a disk and 'blow' it away.

For the $\log(R'_{\rm HK}) - M_\star$ relationship, again absorbing and non-absorbing distributions appear distinct. A similarly low Anderson Darling test scores of 5.37 ($p = 2.57 \times 10^{-3}$) indicates there is a low probability the samples have been drawn from the same distribution. There is a minimum mass threshold of $\sim 0.8 M_\odot$ below which

there is not strong evidence for absorption. This may represent the minimum stellar mass required for the formation of a diffuse disk. There is a less distinct difference between the absorbing and non-absorbing distributions when considering the activity–surface gravity ($\log g$) relationship. Indeed, an Anderson-Darling test score of 2.87 ($p = 0.022$) returns a higher probability that samples have been drawn from the same distribution . This suggests that stellar wind strength, rather than stellar gravity, is the main factor determining disc stability. These findings will help to constrain 3D hydrodynamic models of enshrouded systems.

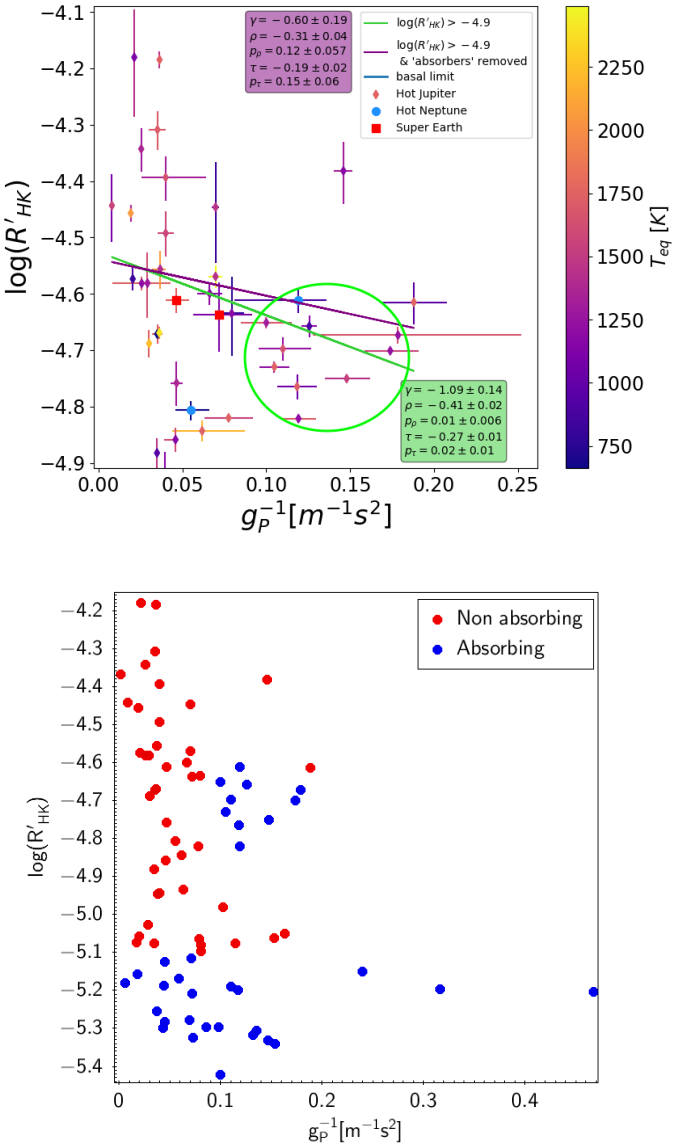

FIGURE 5.20: Top panel: Figure 5.8 replotted with a green circle enclosing 10 high activity sub-samples targets that drive the $\log(R'_{\rm HK}) - g_{\rm P}^{-1}$ anti-correlation. They are high activity systems with low $g_{\rm P}$, which are prone to mass-loss, enshrouding and absorption. Bottom panel: all systems with strong evidence of absorption shown in blue.

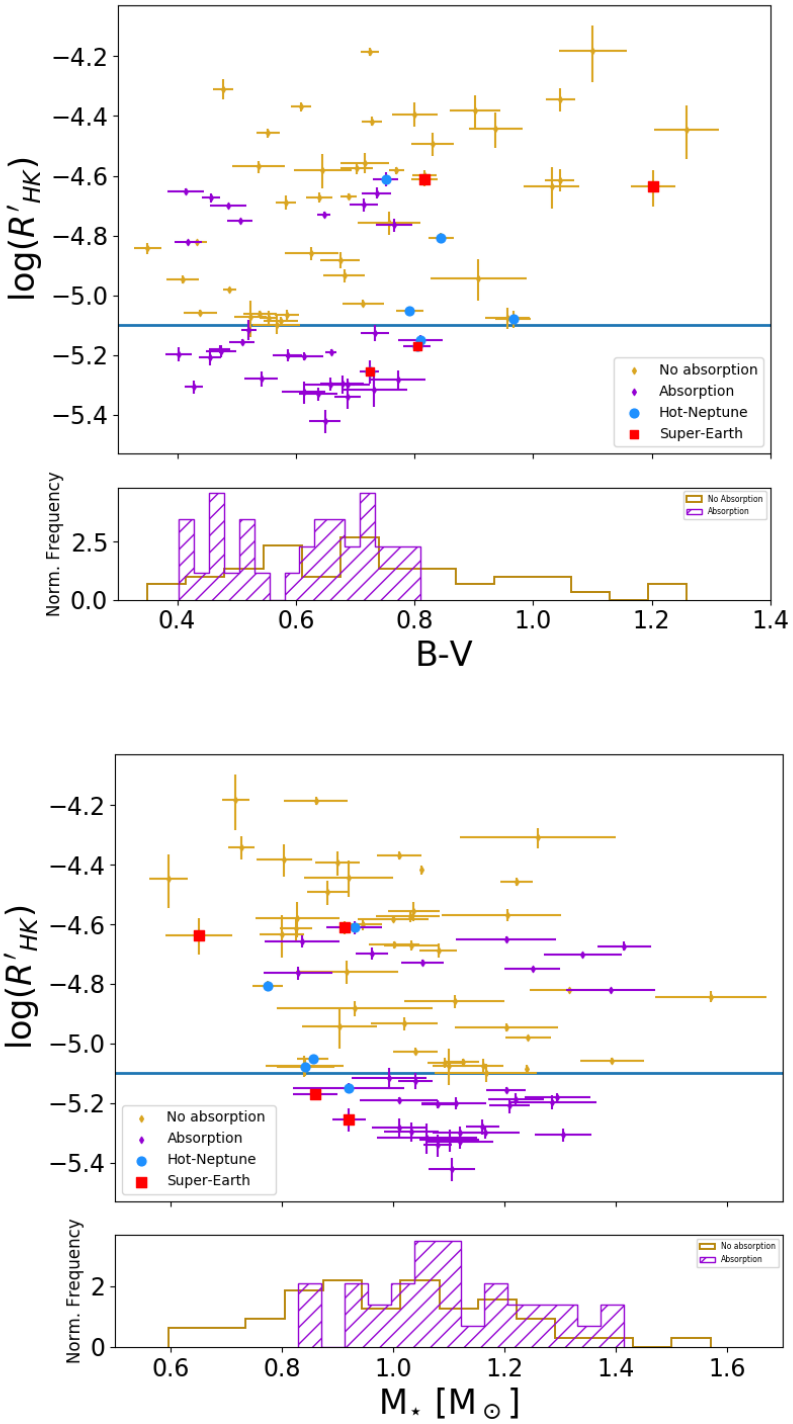

FIGURE 5.21: Colour index (top panel) and stellar mass (bottom panel) plotted against activity for the OU-SALT MS sample. The histograms compare the absorbing and non-absorbing distributions.

5.9 Summary

I have compared the activity distribution of the OU-SALT sample to two activity surveys. I corrected a $S \rightarrow \log(R'_{\rm HK})$ calibration error in the largest activity catalogue compiled to date, which casts doubt on the conclusions of Boro Saikia et al. (2018) (BS18) concerning the weakening of the Vaughan Preston (VP) gap.

The distributions of planet hosts and field stars are distinct: 31% of OU-SALT MS systems display sub-basal activity compared to only 2% of MS field stars, while 75% of OU-SALT SGs have $\log(R'_{\rm HK}) < -5.1$, compared to 29% of SG field stars. These findings comprise strong evidence for mass loss and absorption by diffuse gas disks. I confirmed bimodality in OU-SALT MS sample using Hartigan's dip test value, further indicating the presence of the VP gap. Using tools made available by Redfield & Linsky (2008) and (Fossati et al., 2017b), I have established that absorption by the ISM can not explain the majority of sub-basal values.

A detailed analysis of correlations between planetary/stellar parameters and $\log(R'_{\rm HK})$ was presented. Distinct $\log(R'_{\rm HK}) - g_{\rm P}^{-1}$ trends in the high ($\rho = -0.41 \pm 0.02$, $p = 0.01 \pm 0.006$) and low ($\rho = -0.31 \pm 0.03$, $p = 0.06 \pm 0.026$) activity MS sub-samples indicate that more dramatic mass loss occurs from low-gravity planets around high activity stars. $T_{\rm eq}$ was homogeneously calculated for all planets but $\log(R'_{\rm HK}) - T_{\rm eq}$ correlations are weak ($\rho = 0.06 \pm 0.04$, $p = 0.69 \pm 0.16$).

A significant $\log(R'_{\rm HK}) - a$ correlation in the high activity population ($\rho = -0.29 \pm 0.01$, $p = 0.01 \pm 0.003$) strengthened for systems where $a < 0.05$ AU ($\rho = -0.43 \pm 0.01$, $1.65 \times 10^{-3} \pm 3.64 \times 10^{-4)}$). Similarly, a $\log(R'_{\rm HK}) - M_{\rm P}$ correlation ($\rho = 0.36 \pm 0.03$, $p = 0.02 \pm 0.012$) increased for $M_{\rm P} > 0.8\ M_{\rm J}$ ($\rho = 0.52 \pm 0.01$, $p = 0.01 \pm 0.002$). I have also identified strong correlations between activity and SPI proxies. Close-in, heavy MS planets ($\log M_{\rm P}/a > 0.6$) have a $\log(R'_{\rm HK}) - \log M_{\rm P}/a$ correlation of $\rho = 42 \pm 0.01$ ($p = 5.36 \times 10^{-4} \pm 1.93 \times 10^{-4}$). For 'extreme' MS systems with $M_{\rm P}^2/a > 450\ {\rm M_J\ AU^{-2}}$, $\rho = 0.34 \pm 0.01$ ($p = 2.42 \times 10^{-3}$).

These findings provide tantalising *prima facie* evidence of SPI. However, even with the unique precision of the OU-SALT sample, interdependencies (particularly between $M_\star - M_\mathrm{P}$ and $M_\star - a$) and observational selection effects of transiting hosts must be accounted for before SPI is confirmed. This is the focus of Chapter 6.

Finally, I have placed constraints on the physical parameters of systems in which absorption by diffuse gas disks occurs. Enshrouding happens when $0.4 < B-V < 0.8$ (equating to $5200 \lesssim T_\mathrm{eff} \lesssim 6600$ K) and $M_\star \gtrsim 0.8 M_\odot$. Solar wind strength and radiation pressure influence disk stability more than stellar surface gravity in the mass range of stars we have studied. These results provide important constraints for 3D hydrodynamic models.

In the next chapter, I will attempt to untangle the web of interdependent variables, observational biases, and selection effects to search for meaningful, statistically significant relationships between activity and planetary/stellar variables.

Collaboration details

This work in this chapter was undertaken in regular consultation with Carole Haswell and John Barnes. The findings presented in Section 5.8 followed discussions with Luca Fossati.

Chapter 6

Multivariate analysis of the OU-SALT MS sample

The case for using multivariate methods to study the relationship between planetary/stellar parameters and chromospheric activity is presented in Section 6.1. A multiple linear regression model, which incorporates principal components analysis (PCA) and ordinary least squares regression, is constructed and tested in Section 6.2. The model is optimised by inputting Gaia DR2 variables and employing the 'kernel trick' in Section 6.3. The multivariate model is used to predict the activity of various sub-samples in Section 6.5, and to test the influence of a range of planetary parameters in Section 6.6. A summary and details of collaboration are presented in Section 6.7. In Appendix A, I provide mathematical descriptions of linear and non-linear PCA, kernel functions, and metrics for assessing model prediction success. Appendix B details my attempt to improve model prediction success with supervised learning methods.

Results in this chapter were presented at the European Planetary Sciences Congress—Division for Planetary Sciences Joint Meeting 2019 (Doherty et al., 2019).

6.1 Multivariate statistics

As discussed in Chapter 5, a multitude of variables influence the activity of planet hosting stars. Stellar parameters, including mass, temperature and rotation rate, are directly correlated with magnetic activity and chromospheric emission. The presence of a close-in planet may cause a star to spin up or down, influencing convection in the stellar envelope (Zahn, 2008; Mathis & Le Poncin-Lafitte, 2009; Poppenhaeger, 2016). The companion may also deposit magnetic or gravitational energy in the stellar atmosphere (Cuntz et al., 2000; Shkolnik et al., 2008; Strugarek et al., 2014). These processes intrinsically alter chromospheric activity. At the same time, mass lost from ablating planets may extrinsically alter observed activity (Haswell et al., 2012; Fossati et al., 2013; Staab et al., 2017).

The correlations between $\log(R'_{\rm HK})$ and stellar/planetary parameters identified in Chapter 5 provide non-conclusive indications of star-planet interactions (SPI). However, several authors have highlighted the importance of taking into account selection effects, observation biases, and the interdependency of variables when searching for SPI (Poppenhaeger et al., 2010; Poppenhaeger & Schmitt, 2011; Miller et al., 2015; France et al., 2018). Rigorous statistical treatment is required to investigate mass loss, enshrouding, absorption, and SPI in close-in planetary systems.

Multivariate analysis (MVA) provides the sophisticated statistical tool kit to examine the relationship between multiple dependent (target) and independent (predictor) variables, and to test our hypotheses. It can simultaneously reduce structural complexity in data without losing scientifically significant information. I use the PCA model of France et al. (2018) as a starting point for the modelling undertaken in this chapter, and make extensive use of the *Scikit-learn* Python module (Pedregosa et al., 2011).

6.2 Predicting activity

The following notation is adopted: observations are defined as $\{\mathbf{x}_{nw}, \mathbf{y}_n\}$, where $\mathbf{x}_{nw}$ are the predictor variables (i.e. stellar/planetary parameters), $w = 1, ..., W$ where W is the total number of predictor variables; $\mathbf{y}_n$ is the target variable (i.e. activity), and $n = 1, ..., N$ where N is the total number of stars in the sample. My goal is to build a model that predicts activity values ($\hat{\mathbf{y}}_n$). Uncertainties on predictor and target variables (ϵ_x and ϵ_y respectively) are used to estimate prediction uncertainty ($\epsilon_{\rm pred}$: see Section 6.2.1).

6.2.1 Multiple linear regression model

Linear regression may be extended to multivariate data as follows:

$$\hat{y}_n = \beta_0 + \beta_1 x_{n1} + ... + \beta_W x_{NW}, \tag{6.1}$$

where β_0 is the y-intercept and β_W are the slope coefficients that weight each predictor variable (Feigelson & Babu, 2012). This is a multiple linear regression model: there are multiple predictor variables and a single target variable.

To build a model that predicts chromospheric activity, I first chose suitable predictor variables. Stellar temperature, mass, and rotation period drive activity. I also wanted to account for observational biases and absorption by the interstellar medium (ISM). In a search for SPI signals in far-ultraviolet (FUV; $1150 - 1450$ Å) activity measurements, France et al. (2018) chose the following predictor variables: stellar rotation period ($P_{\rm rot}$), absolute magnitude ($M_{\rm V}$), colour index ($B-V$), distance between the observer and star (d), effective temperature ($T_{\rm eff}$), and the simple SPI term $M_{\rm P}/a$ (where $M_{\rm P}$ is planet mass and a is the orbital semi-major axis).

Taking these variables as a starting point, and assuming (for the moment) that they contribute linearly to chromospheric activity, I constructed the following linear regression model ("M1"):

$$\hat{y}_n \equiv \log(R'_{\rm HK})_{\rm pred:stellar+plan} = \beta_0 + \beta_1(P_{\rm rot}) + \beta_2(V) + \beta_3(B-V) + \beta_4(d) + \beta_5(T_{\rm eff}) + \beta_6(M_P/a). \tag{6.2}$$

Predictions made by M1 may be statistically assessed by comparing them to measured activity. I created a second model ("M2") that is the same as M1 except that it does not include the SPI term:

$$\hat{y}_n \equiv \log(R'_{\rm HK})_{\rm pred:stellar} = \beta_0 + \beta_1(P_{\rm rot}) + \beta_2(V) + \beta_3(B-V) + \beta_4(d) + \beta_5(T_{\rm eff}). \tag{6.3}$$

This model incorporates only stellar predictor variables, and may also be statistically assessed. If M1 better predicts activity than M2, the planet is influencing chromospheric activity. If there is no improvement in predictions, or indeed if M2 makes better predictions, then the planet has no significant influence.

Principal component analysis

Multiple linear regression assumes that predictor variables are not highly correlated but, as established in Chapter 5, some stellar parameters are (e.g. $B-V$ and $T_{\rm eff}$). Principle components analysis (PCA)—a technique widely used in astronomy and machine-learning (Jolliffe, 2002; Gratier et al., 2017; France et al., 2018; Post et al., 2019)—may be exploited to correct for this bias and to generate independent predictor variables. PCA is a dimensionality reduction tool that models the covariance structure (i.e. the relationship between the scatter in variables) of multivariate data. Variance (σ^2) for a sample of N values is:

$$\sigma^2 = \frac{\sum (x_n - \overline{x})^2}{N}, \tag{6.4}$$

where $\overline{x}$ is the sample mean. PCA projects multi-dimensional data onto a sub-space with dimensionality $M < W$ (where $m = 1, ..., M$), while maximizing linear variance (Bishop, 2015).

Eigendecomposition of a predictor variable covariance or correlation matrix finds M eigenvectors ($\mathbf{u}_\mathrm{M}$). Each eigenvector is a principal component (PC) corresponding to M largest eigenvalues λ_M. Each PC is a linear combination of the predictor variables and is orthogonal to, and independent of, each other principal component. For example, $\mathbf{u}_2^\mathrm{T}\mathbf{u}_1 = 0$, where T is the transpose of the eigenvector. PCs that contain information that is poor at explaining variance may be dropped from the analysis. A detailed mathematical description of PCA is provided in Appendix A.

PCA may be used to project OU-SALT predictor variables into a new basis where PCs are orthogonal to each other and may legitimately be used as independent variables in a multiple linear regression model.

Data scaling

PCA searches for maximal variance in data. The range of predictor variables values is of fundamental importance. Some predictor variables—such as distance d, semi-major axis a, and planet mass M_P—have large dynamic ranges. Extreme values could dominate the structure of the data and conceal variations of common lower values. It was therefore necessary to centre and scale predictor variables so that any correlations subsequently identified were not due to differences in scale.

I centred and scaled OU-SALT data in two steps. First, logarithms were taken of the predictor variables with large dynamic ranges: P_rot, d, M_P and a. Applying PCA to the logarithm of data allows a decomposition as a product of factors, and describes the data structure in terms of ratios, products and power laws (Gratier et al., 2017). Next, to address incompatible units and scales in the data, variables were offset to zero mean and rescaled as per Equation 7 of France et al. (2018):

$$\tilde{x}_{nw} = \frac{x_{nd} - \overline{x}_w}{L_x}, \tag{6.5}$$

where L_x is the L2-norm (or Euclidean norm) that calculates the distance of the vector coordinate from the origin of the vector space as follows:

$$L_x = \sqrt{\sum_{n=1}^{N} |x_{nd} - \overline{x}_w|^2}. \tag{6.6}$$

Calculating PCs

Principal components are often calculated using a covariance matrix. However, where input variables have different scales, using a correlation matrix is preferable (e.g. France et al., 2018). Linear correlations were calculated between scaled predictor variables using the Pearson correlation coefficient:

$$Z\left(x_{nj}, x_{nk}\right) = \sum_{n=1}^{N} \tilde{x}_{nj}\tilde{x}_{nk}, \tag{6.7}$$

where the indices j and k represent specific predictor variables (e.g. V and $B-V$). The following correlation matrix was generated for scaled M1 predictor variables.

	$\log(P_{\text{rot}})$	V	$B-V$	T_{eff}	$\log(d)$	$\log(M_P/a)$
$\log(P_{\text{rot}})$	1.00	−0.02	0.52	−0.51	−0.34	−0.24
V	−0.02	1.00	0.30	−0.31	0.54	0.16
$B-V$	0.52	0.30	1.00	−0.99	−0.52	−0.07
T_{eff}	−0.51	−0.31	−0.99	1.00	0.51	0.06
$\log(d)$	−0.34	0.54	−0.52	0.51	1.00	0.24
$\log(M_P/a)$	−0.24	0.16	−0.07	0.06	0.24	1.00

The matrix is symmetric, with perfect correlations between the same predictor variable along the diagonal. The M2 correlation matrix is identical but without $\log(M_P/a)$ correlations. Eigendecomposition was performed on the correlation matrices. PCs were defined by the eigenvectors $\mathbf{u}_W$ of each matrix:

$$\mathbf{Z}\mathbf{u}_W = \lambda_W \mathbf{u}_W. \tag{6.8}$$

Spearman rank coefficients (ρ) between each PC and predictor variable, as well as the explained variance (i.e. information content), of each PC are presented in Figure 6.1. Eigenvectors with the lowest eigenvalues contain the least information. Eigenvalues were ranked highest to lowest to choose the top M eigenvectors to incorporate into further analysis. In descending order, the eigenvalues for the M1 eigenvector matrix were $\lambda_{\mathrm{w,M1}}$ = 2.75, 1.46, 0.94, 0.73, 0.12 and 0.0073. For the M2 eigenvector matrix, $\lambda_{\mathrm{w,M2}}$ = 2.71, 1.36, 0.79, 0.13 and 0.0073. The cumulative explained variance is shown in Figure 6.2. M1, $\mathbf{u}_{1,\mathrm{M1}}$ contained 45.8% of the information; $\mathbf{u}_{2,\mathrm{M1}} - \mathbf{u}_{4,\mathrm{M1}}$ contained 24.4%, 15.6% and 12.1% respectively; $\mathbf{u}_{5,\mathrm{M1}}$ and $\mathbf{u}_{6,\mathrm{M1}}$ together accounted for 2.1% of variance. For Model 2, $\mathbf{u}_{1,\mathrm{M2}} = 54.3\%$, $\mathbf{u}_{2,\mathrm{M2}} = 27.2\%$ and $\mathbf{u}_{3,\mathrm{M2}} = 15.8\%$; $\mathbf{u}_{4,\mathrm{M2}}$ accounted for 2.6% of variance; $\mathbf{u}_{5,\mathrm{M2}}$ accounted for only 0.1% of variance.

A PC may be dropped from the analysis if it is not "strongly correlated" ($\rho > 0.5$: Peres-Neto et al., 2005) with any predictor variable. Figure 6.1 shows that the first PC is strongly correlated with P_{rot}, $B-V$, d, and T_{eff}, and the second with M_{V} and T_{eff} for both models. The third PC, $\mathbf{u}_{3,\mathrm{M1}}$, is strongly correlated with $\log(M_{\mathrm{P}}/a)$, and $\mathbf{u}_{3,\mathrm{M2}}$ with P_{rot}. The fourth PC, $\mathbf{u}_4$, has correlations with P_{rot} and $\log(M_{\mathrm{P}}/a)$ that just exceed $\rho = 0.5$ for M1. While $B-V$ and d correlations are present for M2, $\mathbf{u}_{4,\mathrm{M2}}$ explains 9.5% less variance than $\mathbf{u}_{4,\mathrm{M1}}$. Neither $\mathbf{u}_5$ or $\mathbf{u}_6$ are strongly correlated with predictor variables. As such, $\mathbf{u}_{5,\mathrm{M1}}$, $\mathbf{u}_{6,\mathrm{M1}}$, $\mathbf{u}_{4,\mathrm{M2}}$ and $\mathbf{u}_{5,\mathrm{M2}}$ were dropped from the analysis.

Linear model with PC input variables

Incorporating the chosen PCs, M1 and M2 may be described as:

$$\begin{aligned}\log(R'_{\mathrm{HK}})_{\mathrm{pred:stellar+plan}} = \beta_0 + [\beta_{\mathbf{u}1,\mathrm{M1}} \times \mathbf{u}_{1,\mathrm{M1}}] + [\beta_{\mathbf{u}2,\mathrm{M1}} \times \mathbf{u}_{2,\mathrm{M1}}] \\ + [\beta_{\mathbf{u}3,\mathrm{M1}} \times \mathbf{u}_{3,\mathrm{M1}}] + [\beta_{\mathbf{u}4,\mathrm{M1}} \times \mathbf{u}_{4,\mathrm{M1}}];\ \text{and}\end{aligned} \tag{6.9}$$

$$\begin{aligned}\log(R'_{\mathrm{HK}})_{\mathrm{pred:stellar}} = \beta_0 + [\beta_{\mathbf{u}1,\mathrm{M2}} \times \mathbf{u}_{1,\mathrm{M2}}] + [\beta_{\mathbf{u}2,\mathrm{M2}} \times \mathbf{u}_{2,\mathrm{M2}}] \\ + [\beta_{\mathbf{u}3,\mathrm{M2}} \times \mathbf{u}_{3,\mathrm{M2}}].\end{aligned} \tag{6.10}$$

FIGURE 6.1: Bar charts showing the Spearmans rank coefficient value between each predictor variable and principal component (PC: $\mathbf{u}_{\mathrm{M}}$). The percentage of total information contained in each PC is stated.

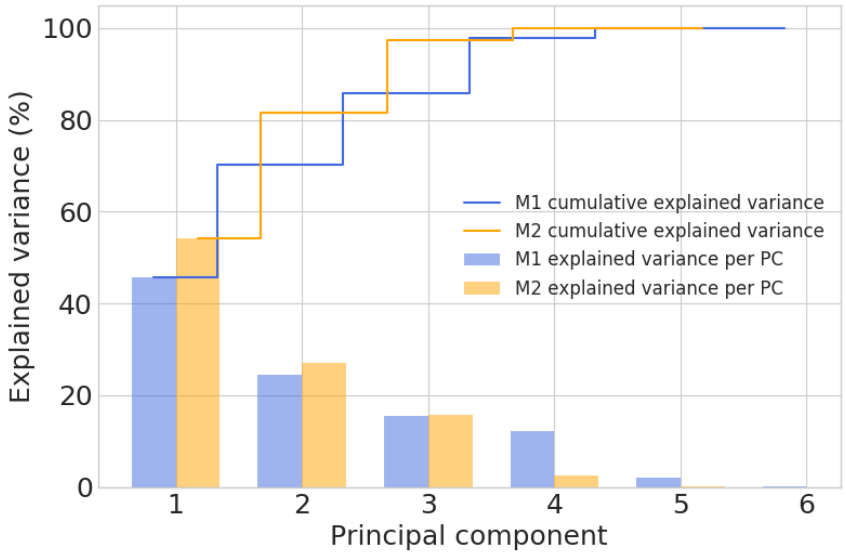

FIGURE 6.2: Explained variance ratio for each PC for the OU-SALT MS sample.

The β coefficients—estimated using ordinary least squares (OLS) regression—and corresponding confidence intervals, are presented in Table 6.1. PCA was carried out separately for M1 and M2, and the SPI term was not included in the PCA or linear regression of M2. The PC inputs and β coefficients for M1 and M2 are distinct. I call this the 'OLS-PCA ' model.

Model	Principal component	β	95% confidence interval
M1	$\mathbf{u}_{1,\mathrm{M1}}$	0.1192	$(0.009, 0.229)$
	$\mathbf{u}_{2,\mathrm{M1}}$	0.3608	$(0.210, 0.511)$
	$\mathbf{u}_{3,\mathrm{M1}}$	-0.3645	$(-0.553, -0.176)$
	$\mathbf{u}_{4,\mathrm{M1}}$	-0.2257	$(-0.439 - 0.012)$
	y-intercept	1.353e-16	$(-0.021, 0.021)$
M2	$\mathbf{u}_{1,\mathrm{M2}}$	0.1433	$(0.027, 0.260)$
	$\mathbf{u}_{2,\mathrm{M2}}$	-0.2680	$(-0.433, -0.103)$
	$\mathbf{u}_{3,\mathrm{M2}}$	-0.4490	$(-0.665, -0.233)$
	y-intercept	1.353e-16	$(-0.022, 0.022)$

TABLE 6.1: Multiple linear regression coefficients with corresponding confidence intervals (calculated using Students t-test) for each term in M1 (Equation 6.9) and M2 (Equation 6.10).

Estimating prediction uncertainty

Propagating errors through complex multivariate analyses by conventional methods is not trivial. (This is particularly the case when using non-linear methods, such as the kernel technique described in Section 6.3.2, or complex machine-learning algorithms: Appendix B). I use a Monte Carlo-type approach, known as the perturbation technique (Curran, 2014), to estimate uncertainties on predictions. This involves creating multiple new data sets where x_{nw} and y_n are perturbed from their measured value by a number drawn randomly from a Gaussian distribution of standard deviation equal to the measured errors on predictor and target variables:

$$\begin{aligned} x_{nw} &= x_{nw} + \mathcal{G}(\epsilon_x), \\ y_n &= y_n + \mathcal{G}(\epsilon_y). \end{aligned} \tag{6.11}$$

The standard deviation of 1000 perturbed values corresponding to each measurement provides 1 σ uncertainty estimates for each prediction.

6.2.2 OLS-PCA results

Predictions generated by M1 (Equation 6.9) and M2 (Equation 6.10) are plotted against scaled observed $\log(R'_{\mathrm{HK}})$ values in Figure 6.3(a). M1 and M2 predictions track each other closely, with several exceptions, and are scattered about the plotted line of equivalence. Figure 6.3(b), which shows residuals between measured and predicted activity ($\hat{e}_n = y_n - \hat{y}_n$) for M1 and M2, was used to check for discrepancies and outliers (an approach advocated by Kutner, 2005; Sheather, 2009; Feigelson & Babu, 2012). Figure 6.4 shows the distributions of these residuals, which are well-centred but with peaks either side of zero. The standard deviations of the distributions (σ_{res}) calculated as a percentage of the standard deviation of measured activity are 77% and 82% respectively. This indicates that M1 finds a better fit for the data than M2.

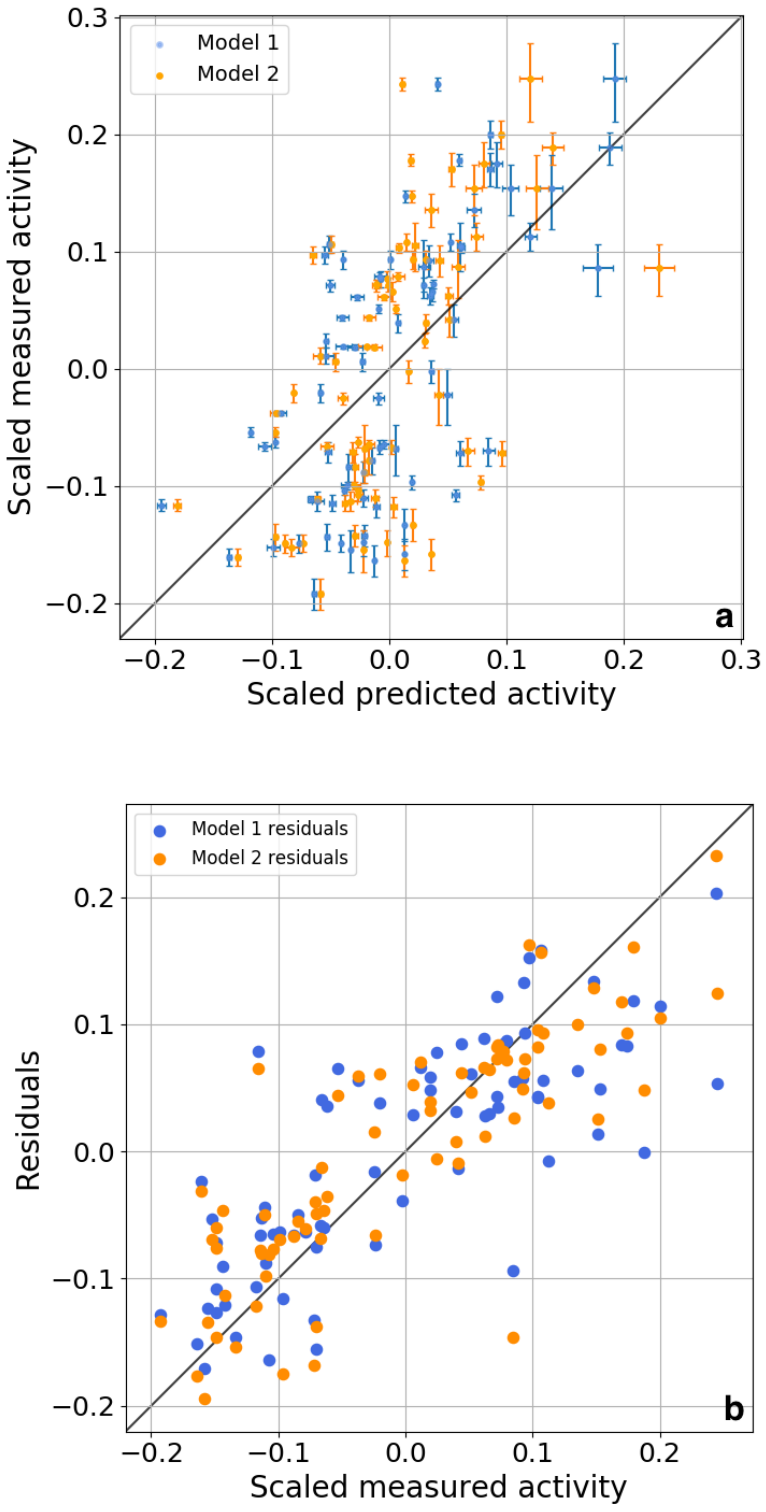

FIGURE 6.3: (a) OLS-PCA predictions with 1 σ uncertainties for M1 (blue) and M2 (orange). (b) Residuals ($\hat{e}_n = y_n - \hat{y}_n$) plotted against scaled measured activity.

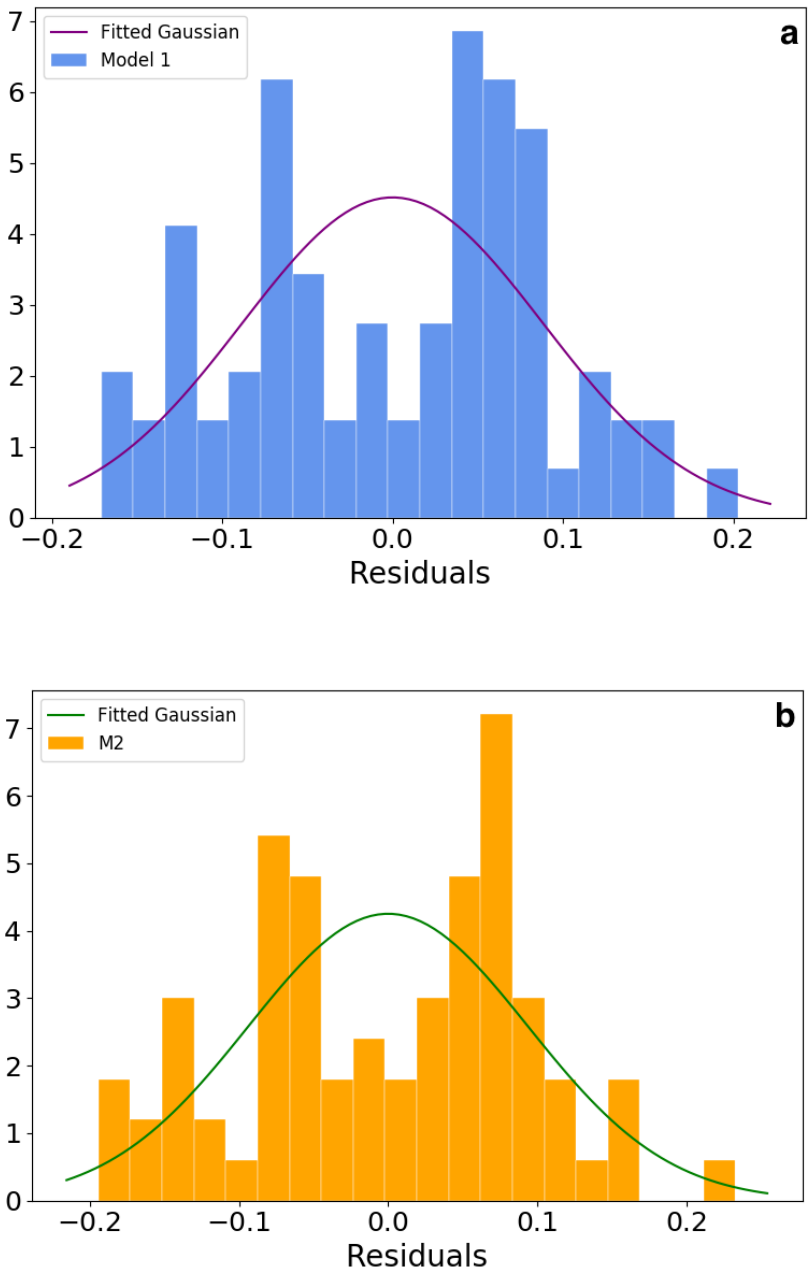

FIGURE 6.4: Residual distributions for (a) M1 and (b) M2. Residual values are shown on the x-axis and frequency on the y-axis.

Assessing prediction power

To evaluate how successfully each model is able to predict activity, I employed three Monte Carlo based methods: bootstrap, permutation testing and the perturbation method.

Bootstrap resampling involves the construction of hypothetical datasets derived from observations and predictions, each of which may be statistically assessed (Efron, 1979; Singh, 1981). The datasets—which are the same size as the original set—are constructed using 'resampling by replacement'. Each point is selected from the full dataset. Predictions may appear more than once in each bootstrapped sample or not at all. (This is also known as 'out-of-the-bag' sampling.) Bootstrapped distributions reflect the variance in the data and may be used to quantify uncertainties, confidence intervals and success metrics. Bootstrap does not assume that data are normally distributed.

I generated 1000 bootstrapped samples from M1 and M2 predictions, and statistically evaluated each using the following success metrics: root-mean-square error (RMSE), adjusted R^2 (R^2_a), the Bayesian Information Criterion (BIC), and σ_{res} (mathematical descriptions of which are provided in Appendix A.4). I also quantified the correlation between predicted and observed activity in each sample with the Spearman's rank coefficient. RMSE and R^2_a indicate how much scatter a model explains. R^2_a penalizes the number of parameters used, so may indicate if a model is overfitting data. The BIC contains a penalty term that scales with both the number of parameters and sample size. BIC values indicate how useful a model is for making predictions.

Success metric distributions for OLS-PCA predictions are shown in Figure 6.5. Bootstrapped distributions are outlined in black and red. The mean and standard deviation of each distribution is presented in Table 6.2. M1 outperformed M2 in all instances with the following differences between metrics: $\Delta\,\rho = 0.01$, $\Delta\,R^2_a = 0.06$, Δ RMSE = -0.005, Δ BIC = -5 and $\Delta\,\sigma = -5\%$.

Bootstrap has limitations. In particular, it assumes a sample is representative of the overall population, which may be problematic for smaller samples (Haukoos & Lewis, 2005). As such, the perturbation method was also employed to estimate uncertainty. Perturbed distributions are shown in blue and orange in Figure 6.5. Again, M1 outperformed M2 with $\Delta\rho = 0.02$, $\Delta R^2_a = 0.08$, Δ RMSE = -0.006, Δ BIC = -9 and $\Delta\sigma = -5\%$. It is notable that the standard deviation of distributions generated using the perturbation method are considerably smaller than those of the bootstrap method.

	ρ	R^2_a	RMSE	BIC	$\sigma_{res}[\%]$
Bootstrap					
M1					
μ_{BS1}	0.59	0.34	0.088	-362	77
σ_{BS1}	0.084	0.084	0.005	9.7	4.82
M2					
μ_{BS2}	0.58	0.28	0.093	-357	82
σ_{BS2}	0.090	0.088	0.006	10.2	5.43
S_{1-2}	0.07	0.55	-0.64	-0.33	-0.63
p_{1-2}	0.94	0.58	0.52	0.74	0.53
Perturbation method					
M1					
μ_{PM1}	0.60	0.36	0.088	-361	78
σ_{PM1}	0.017	0.034	0.002	4.2	2.08
M2					
μ_{PM2}	0.58	0.28	0.094	-352	83
σ_{PM2}	0.019	0.029	0.002	3.1	1.66
S_{1-2}	0.76	1.84	-1.83	-1.82	-1.83
p_{1-2}	0.45	0.07	0.07	0.07	0.07
Permutation test					
μ_n	0.000	-0.47	0.134	-297	118
σ_n	0.111	0.146	0.007	7.8	5.87
M1					
$S_{1\text{-}n}$	4.24	5.11	-5.29	-5.23	-5.35
$p_{1\text{-}n}$	2.25e-05	3.30e-07	1.24e-07	1.74e-07	8.72e-08
M2					
$S_{2\text{-}n}$	4.08	4.60	-4.46	-4.69	-4.51
$p_{2\text{-}n}$	4.43e-05	4.23e-06	8.02e-06	2.72e-06	6.57e-06

TABLE 6.2: Success metrics (ρ, R^2_a, RMSE, BIC and σ_{res}) for the OLS-PCA model. The μ_{BS1} and μ_{BS2} values are the mean of a distribution of 1000 samples bootstrapped from M1 and M2 predictions, with associated standard deviations (σ_{BS1} and σ_{BS2}). The μ_{PM1} and μ_{PM1} values are means from 1000 perturbed M1 and M2 predictions, with associated standard deviations (σ_{PM1} and σ_{PM2}). The μ_n value is the mean of the null permuted distribution. Separations between M1 and M2 distributions are stated, along with associated p values. R^2_a and BIC are accorded most weight as they penalise additional predictor variables.

Significance

The separation S_{1-2} between the distributions may be calculated:

$$S_{1-2} = \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}, \tag{6.12}$$

where μ_1 and μ_2 are the means of the M1 and M2 success metric distributions respectively, and σ_1 and σ_2 are the respective standard deviations. S_{1-2} describes how many σ separate the M1 and M2 distributions and, in so doing, quantifies statistical significance. Table 6.2 shows $|S_{1-2}|$ separations of $0.07 - 0.64\,\sigma$ for bootstrapped distributions, and $0.76 - 1.84\,\sigma$ for perturbed distributions. It is noteworthy that the success metrics (excluding ρ)—all of which have different mathematical underpinnings—show similar statistical significance. S_{1-2} values are also translated into p-values which, for the perturbed method, indicate a 7 % chance that the model with the SPI parameter explains the same variance as the model without.

I also employed permutation testing for hypothesis testing. This involves calculating many values of the test statistics under rearrangements of predicted data points (Jowett & Fisher, 1956). I decoupled predictions from target observations, then shuffled the predicted data points 1000 times. Unlike bootstrap sampling, this is not done by replacement but involves a shuffle of the entire dataset. Each permuted sample contains all predicted values and may be used to test data against the null hypothesis. Table 6.2 shows conclusive results: $|S_{1-n}|$ and $|S_{2-n}|$ values are more than $5\,\sigma$ for M1 and $4\,\sigma$ for M2. There is a very low probability ($p = 2.25 \times 10^{-5} - 8.72 \times 10^{-8}$) that the variance explained by the models is down to chance.

The models explain $\sim$ 36% (M1) and $\sim$ 28% (M2) of variance in the data respectively. If possible, it is desirable to improve the prediction power of the models.

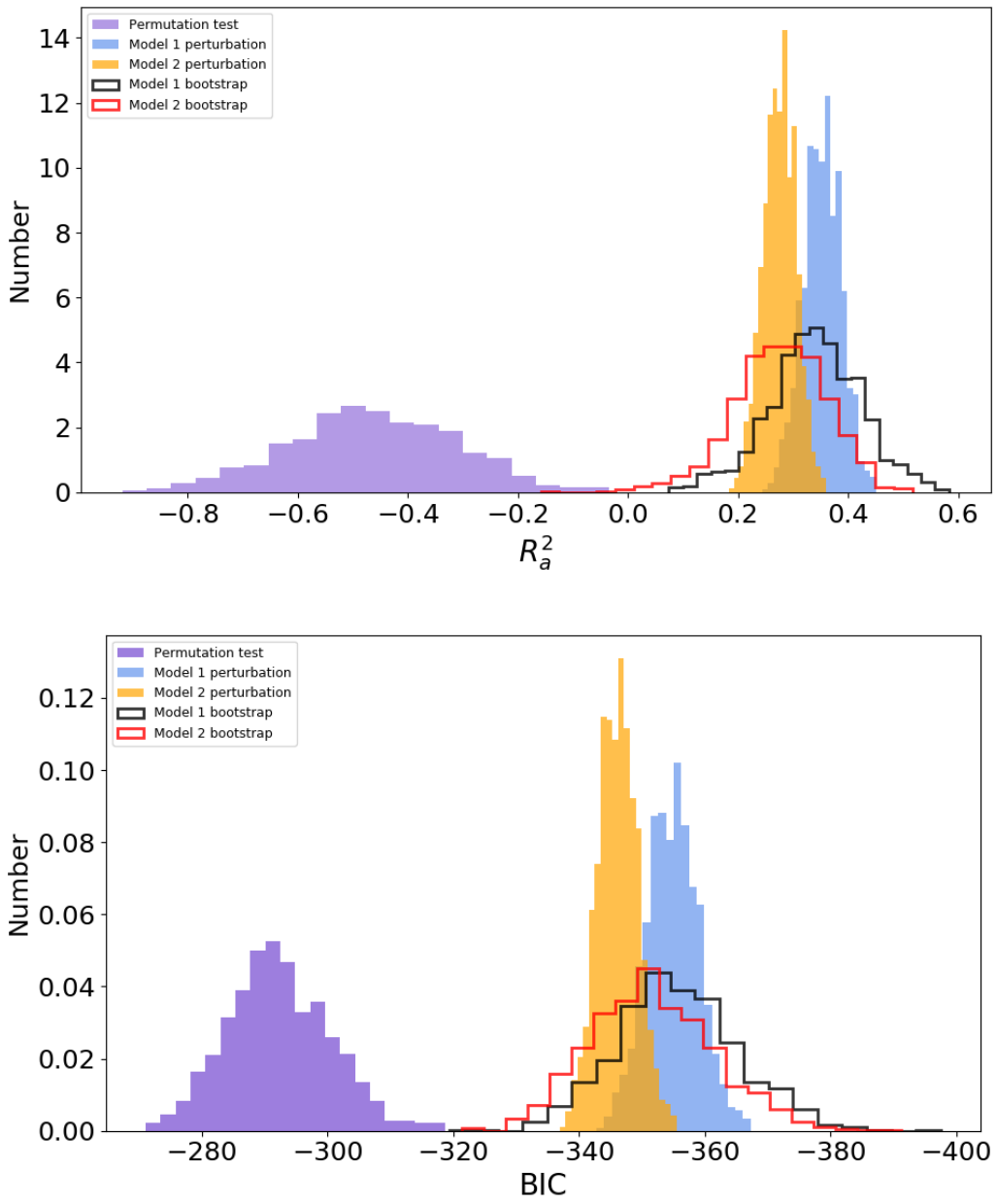

FIGURE 6.5: R_a^2 and BIC distributions for OLS-PCA Models 1 & 2. Each metric is presented in a separate panel. Larger R_a^2, and larger negative BIC values, indicate a stronger fit. Bootstrapped distributions are outlined in black and red, permuted distributions are filled purple, and perturbed distributions are filled blue and orange (see legends). Histogram counts are normalized.

6.3 Linear model optimization

Three improvements to the OLS-PCA model were trialled: using Gaia measurements; incorporating non-linear PCA; and optimizing input variables.

6.3.1 Gaia DR2 data

With the exception of distance d, predictor variables have been measured by a range of instruments. Systematic offsets between instruments introduce uncertainty and scatter. The Gaia DR2 database contains absolute magnitude (G_V), colour index ($B_G - R_G$), and effective temperatures ($T_{\rm eff,G}$) measurements that have been made with a single instrument and so should contain minimal systematic offset. Substituting $B-V$, V and $T_{\rm eff}$ with $R_{\rm G} - B_{\rm G}$, $G_{\rm V}$ and $T_{\rm eff,G}$ in M1 and M2 yields the results shown in Table 6.3. M1 accounted for 2% more variance and M2 for 3% more variance. RMSE values stayed relatively stable, $\sigma_{\rm res}$ remained the same, and the BIC value improved by 3 for M1 and 4 for M2. Gaia predictor variables were therefore used for further modelling, with WASP-103 (noted in Section 4.2.1 to have anomalously large errors in Gaia DR2 measurements) removed from the input data.

	ρ	$R^2_{\rm a}$	RMSE	BIC	$\sigma_{\rm res}$ [%]
Non-Gaia inputs					
M1					
$\mu_{\rm BS1}$	0.59	0.34	0.088	-362	77
$\sigma_{\rm BS1}$	0.084	0.084	0.005	9.7	4.82
M2					
$\mu_{\rm BS2}$	0.58	0.28	0.093	-357	82
$\sigma_{\rm BS2}$	0.090	0.088	0.006	10.2	5.43
Gaia inputs					
M1					
$\mu_{\rm BS1}$	0.59	0.36	0.087	-365	76
$\sigma_{\rm BS1}$	0.088	0.084	0.005	9.9	4.83
M2					
$\mu_{\rm BS2}$	0.58	0.31	0.091	-361	80
$\sigma_{\rm BS2}$	0.090	0.093	0.006	10.5	5.44

TABLE 6.3: Success metrics for bootstrapped distributions when the OLS-PCA model is re-run with Gaia DR2 measurements $R_G - B_G$, G_V and $T_{\rm eff,\,G}$ substituted for $B-V$, V and $T_{\rm eff}$ predictor variables.

6.3.2 Non-linear PCA

The OLS-PCA model assumes a linear correlation between target and predictor variables. To account for non-linearity, and improve model predictions, I used the 'kernel trick'. A kernel function computes the inner products between all pairs of data, projecting these values into a high dimensional feature space. Data that is not linearly separable in W dimensional space may be so divisible in a higher M dimensional space. The 'kernel trick' involves replacing scalar products in algorithms by a kernel function, which is given by the relation:

$$k(\mathbf{x}, \mathbf{x}') = \phi(\mathbf{x})^T \phi(\mathbf{x}'), \tag{6.13}$$

where $\mathbf{x}$ and $\mathbf{x}'$ are two samples. The kernel function $\phi(\mathbf{x})$ is non-linear. It maps the original variables into a larger M dimensional feature space by creating non-linear combinations of the original features (Aizerman et al., 1964; Boser et al., 1992). The kernel is a symmetric function so $k(\mathbf{x}, \mathbf{x}') = k(\mathbf{x}', \mathbf{x})$.

I incorporated the kernel trick into my PCA by applying a non-linear transformation $\phi(x)$ to project each data point onto $\mathbf{x}_{nw} \rightarrow \phi\mathbf{x}_{nw}$ in an M-dimensional feature space. Linear decision boundaries were applied in the higher-dimensional space to define non-linear PCs, then translated back to lower dimensional space. This method is known as non-linear or kernel PCA (kPCA: Schölkopf et al., 1997). A detailed mathematical description of this technique is provided in Appendix A.2.

The following kernel functions—each of which is described in Appendix A.3—were trialled with the OLS model: Polynomial, Radial Basis Function (RBF), Sigmoid and Cosine. The results are presented in Table 6.4. Polynomial, RBF and Sigmoid kernel functions did not significantly improve the models. However, the OLS model with a cosine kernel explained 7 % more variance for M1, increased the correlation by 0.07, dropped residual distribution width by 4 %, and improved the BIC score by 9. M2 prediction success also improved. The cosine kernel was incorporated into the OLS models.

Table 6.4 also describes the performance of the models when the number of PC inputs are varied. Maximum variance is explained by M1 (50 %) and M2 (39 %) when 6 PCs and 5 PCs are used respectively.

	ρ	R^2_a	RMSE	BIC	σ_{res}[%]
Linear					
M1	0.59	0.36	0.087	-365	76
M2	0.58	0.31	0.091	-361	80
Poly					
M1	0.59	0.36	0.087	-365	76
M2	0.58	0.31	0.091	-361	80
RBF					
M1	0.59	0.37	0.086	-365	76
M2	0.59	0.31	0.091	-362	80
Sigmoid					
M1	0.59	0.36	0.087	-364	76
M2	0.58	0.31	0.091	-361	80
Cosine					
M1	0.66	0.43	0.082	-374	72
M2	0.58	0.32	0.090	-363	79
3 PCs					
M1	0.58	0.35	0.088	-366	77
M2	0.58	0.32	0.090	-363	79
4 PCs					
M1	0.66	0.43	0.082	-374	72
M2	0.59	0.33	0.089	-360	7
5 PCs					
M1	0.68	0.44	0.081	-372	71
M2	0.64	0.39	0.085	-364	74
6 PCs					
M1	0.71	0.50	0.076	-377	66
M2	0.64	0.38	0.084	-360	74

TABLE 6.4: Success metrics where each of five kernel functions—Linear, Poly, RBF, Sigmoid and Cosine—is incorporated into the OLS-kPCA model. M1 and M2 use 4 and 3 PCs respectively. Beneath the horizontal line: how model success changes with the number of PC inputs (with a cosine kernel is incorporated).

6.3.3 Predictor variable optimization

To test whether the inclusion of each variable was justified, the OLS-kPCA models were run with and without each predictor variable. The results are presented in Table 6.5. Removing d allowed M1 and M2 to explain 1 % and 2 % more variance respectively, while BIC values strengthened by 3. This provides further reassurance that absorption in the ISM has not significantly influenced our results (see Section 5.3). Distance d was dropped from further analysis.

Removal of the other predictor variables weakened prediction success. Despite $R_G - B_G$ and $T_{\mathrm{eff,\,G}}$ being highly correlated, removing either variable led to considerably weaker models. In contrast, removing T_{eff} from the OLS-PCA models improved prediction power. This confirms that kPCA imposes different decision boundaries to linear PCA that identify non-linear variance. The addition of stellar mass ($M_\star$), radius ($R_\star$), metallicity [Fe/H] and density ($\log g$) as input variables did not improve prediction success.

	M1					**M2**				
Removed	ρ	R^2_{a}	RMSE	BIC	σ_{res}[%]	ρ	R^2_{a}	RMSE	BIC	σ_{res}[%]
None	0.71	0.50	0.076	-377	66	0.64	0.39	0.085	-364	74
$\log(P_{\mathrm{rot}})$	0.71	0.47	0.078	-373	68	0.61	0.32	0.089	-356	78
$R_G - B_G$	0.68	0.45	0.079	-370	70	0.63	0.37	0.085	-362	75
G_V	0.67	0.45	0.079	-371	69	0.60	0.32	0.089	-356	78
$\log(d)$	0.72	0.51	0.074	-380	65	0.66	0.41	0.083	-367	73
$T_{\mathrm{eff\,G}}$	0.67	0.43	0.081	-367	71	0.61	0.34	0.087	-359	77
Added										
$\log M_\star$	0.68	0.44	0.080	-368	70	0.58	0.32	0.089	-356	78
$\log R_\star$	0.70	0.48	0.077	-374	68	0.58	0.32	0.089	-356	78
[Fe/H]	0.66	0.42	0.081	-366	71	0.58	0.31	0.090	-355	79
$\log g$	0.69	0.45	0.079	-370	70	0.59	0.33	0.088	-357	77

TABLE 6.5: Success metrics when each stellar predictor variable is removed from the OLS-kPCA model. Note, the top row shows baseline results when all five stellar variables are included. In the lower half, the addition of several other stellar variables—mass ($M_\star$), radius ($R_\star$), metallicity ([Fe/H]), and density ($\log g$)—to the OLS-kPCA model are also tested. These variables do not improve prediction success.

6.4 OLS-kPCA model

The OLS-kPCA M1 and M2 models are described as follows:

$$\begin{aligned}\log(R'_{\rm HK})_{\rm pred:stellar+plan} = \beta_0 &+ [\beta_{\mathbf{u}1,\rm M1} \times \mathbf{u}_{1,\rm M1}] + [\beta_{\mathbf{u}2,\rm M1} \times \mathbf{u}_{2,\rm M1}] \\ &+ [\beta_{\mathbf{u}3,\rm M1} \times \mathbf{u}_{3,\rm M1}] + [\beta_{\mathbf{u}4,\rm M1} \times \mathbf{u}_{4,\rm M1}] \\ &+ [\beta_{\mathbf{u}5,\rm M1} \times \mathbf{u}_{5,\rm M1}];\end{aligned} \tag{6.14}$$

$$\begin{aligned}\log(R'_{\rm HK})_{\rm pred:stellar} = \beta_0 &+ [\beta_{\mathbf{u}1,\rm M2} \times \mathbf{u}_{1,\rm M2}] + [\beta_{\mathbf{u}2,\rm M2} \times \mathbf{u}_{2,\rm M2}] \\ &+ [\beta_{\mathbf{u}3,\rm M2} \times \mathbf{u}_{3,\rm M2}] + [\beta_{\mathbf{u}4,\rm M2} \times \mathbf{u}_{4,\rm M2}].\end{aligned} \tag{6.15}$$

A principle component was dropped from both M1 and M2 as d was no longer an input variable. The β coefficients are shown in Table 6.6. Figure 6.6 shows correlations between kernel PCs and predictor variables. The first PC is strongly correlated with stellar variables. The second and third PCs are correlated with $\log(P_{\rm rot})$ and the SPI term. The fourth and fifth principal components do not demonstrate strong correlations with predictor variables but do nevertheless explain variance and improve predictions.

Model	Kernel PC	β	95% confidence interval
M1	$\mathbf{u}_{1,\rm M1}$	0.0383	$(0.014, 0.063)$
	$\mathbf{u}_{2,\rm M1}$	-0.1253	$(-0.159, -0.092)$
	$\mathbf{u}_{3,\rm M1}$	0.0601	$(0.017, 0.103$
	$\mathbf{u}_{4,\rm M1}$	-0.1903	$(-0.317, -0.064)$
	$\mathbf{u}_{5,\rm M5}$	-0.3307	$(-0.536, -0.126)$
	y-intercept	-6.852e$-$17	$(-0.018, 0.018)$
M2	$\mathbf{u}_{1,\rm M2}$	0.0423	$(0.018, 0.067)$
	$\mathbf{u}_{2,\rm M2}$	0.1087	$(0.070, 0.147)$
	$\mathbf{u}_{3,\rm M2}$	0.1229	$(0.030, 0.215)$
	$\mathbf{u}_{4,\rm M2}$	0.2597	$(0.082, 0.437)$
	y-intercept	1.353e$-$16	$(-0.020, 0.020)$

TABLE 6.6: Multiple linear regression coefficients with corresponding confidence intervals (calculated using Students t-test) for each term in Equation 6.14 and Equation 6.15.

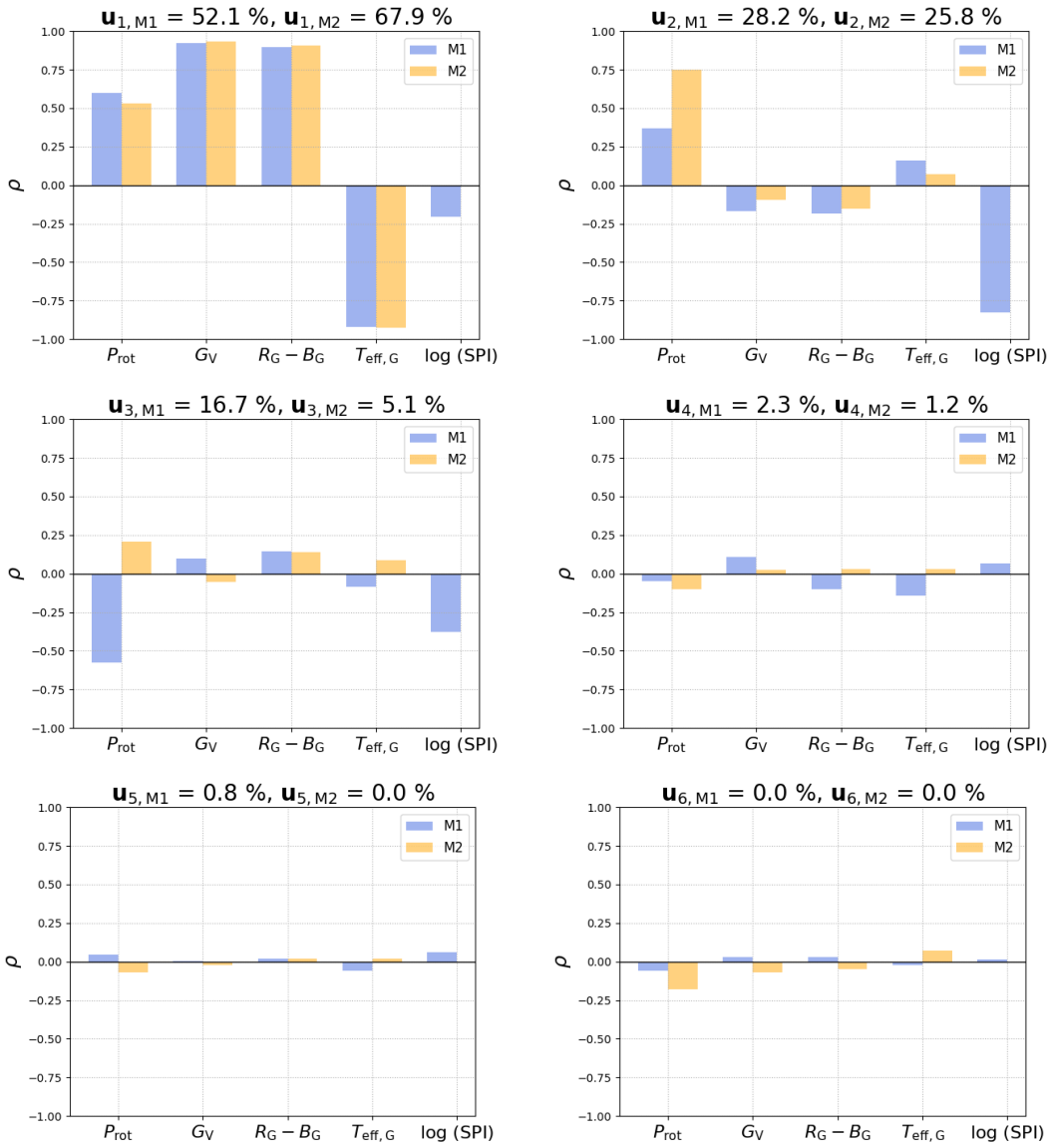

FIGURE 6.6: Bar charts showing the Spearmans rank coefficient value between each predictor variable and kernel PC ($\mathbf{u}_{\mathrm{M}}$). Percentage of total information contained in each kernel PC stated.

6.4.1 Results

Applying the optimized OLS-kPCA model to the OU-SALT MS generates the predictions shown in Figure 6.7 (panel c). M1 and M2 residuals ($y_n - \hat{y}_n$) are also shown (in panel d). Figure 6.8 (panels c and d) shows histograms of the residual distributions for M1 and M2, along with fitted Gaussian curves. The residual distribution widths have decreased by 12 % and 9 % from the OLS-PCA model, indicating a marked improvement in performance. The fitted Gaussian curves are well centred ($\mu = 1.08 \times 10^{-18}$ and 2.16×10^{-18} respectively), indicating an absence of bias.

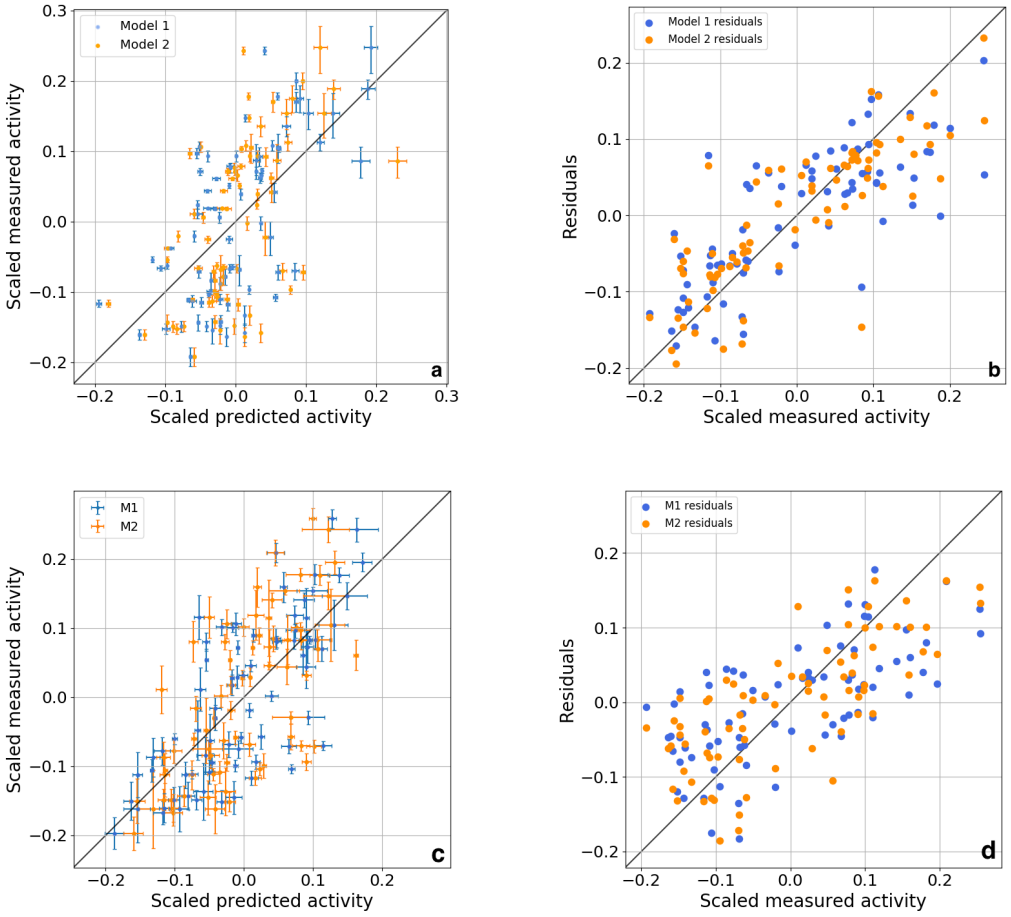

FIGURE 6.7: (a) OLS-PCA: predictions with 1 σ error bars plotted against scaled measured activity (b) OLS-PCA: residual between predicted and measured activity (c) OLS-kPCA: predictions with 1 σ error bars plotted against scaled measured activity (d) OLS-kPCA: residual between predicted and measured activity.

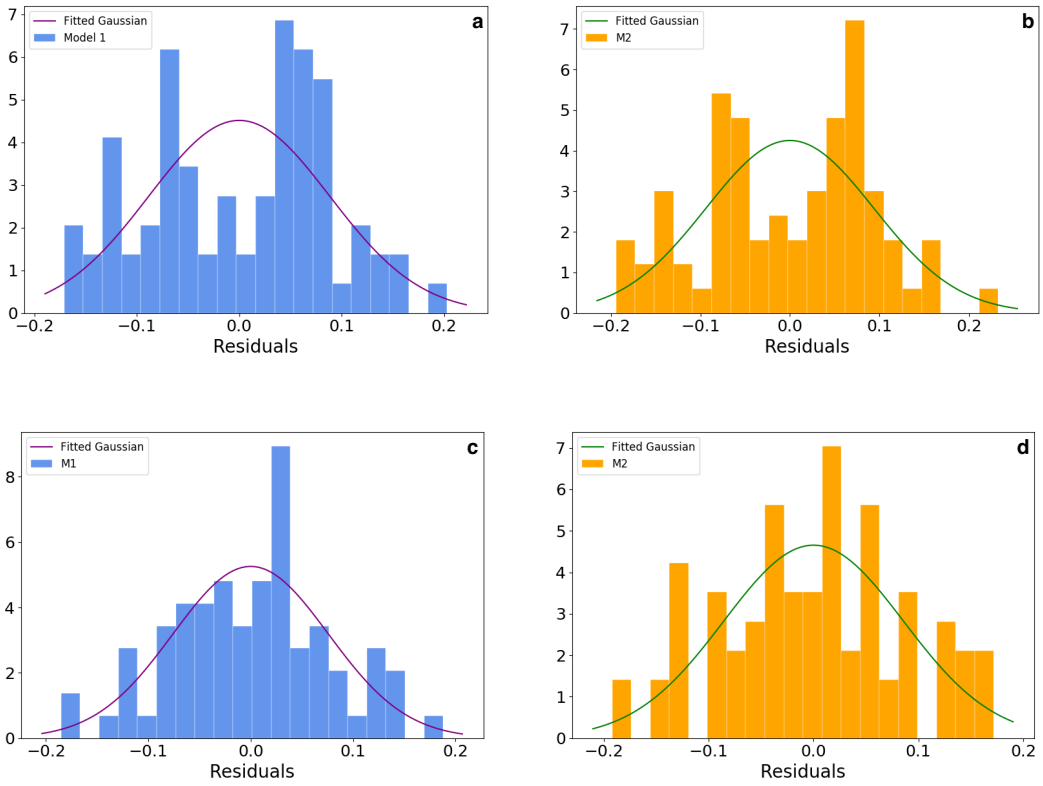

FIGURE 6.8: Residuals of predicted and measured activity for: (a) OLS-PCA M1; (b) OLS-PCA M2; (c) OLS-kPCA M1; and (d) OLS-kPCA M2. Residual values are shown on the x-axis and frequency on the y-axis.

Results for the OLS-kPCA models are shown in Table 6.7. Despite having one less predictor variable than the OLS-PCA model, the OLS-kPCA model explains 13 % more variance for M1 and 8 % more variance for M2. For the bootstrapped samples, the separation between models is between $0.68 - 1.10\,\sigma$. Separation is larger for the perturbed distributions, with $|S_{1-2}| = 2.02 - 2.64\,\sigma$. This indicates that there is only a $0.8 - 0.9$ % likelihood that the models account for the same variance. Figure 6.9 shows success metric distributions. Again, the separation between the perturbed distributions is larger than between bootstrapped distributions.

To gain greater statistical precision and account for uncertainties identified by both the bootstrap and perturbation methods, I have combined these approaches. Each perturbed data set was bootstrapped 100 times, with mean success metrics collected for each sample. This generated 100 000 samples, each with associated success metrics. Figure 6.9 shows that the combined distributions resemble the perturbed distributions but shift towards the bootstrapped mean. Most notably, the mean M2 BIC score shifts by 10. Separation values, which range between $|S_{1-2}| = 1.96 - 2.56\,\sigma$, are more than for bootstrapped but less than for perturbed distributions.

The OLS-kPCA model results indicate that predictions improve with the incorporation of the $\log(M_{\rm P}/a)$ term, with up to 2.64 σ ($p = 0.008$) significance. An attempt to further improve prediction power using supervised learning is described in Appendix B.

Stellar rotation period

M2 should contain minimal planet-related information. As noted in Section 5.7 however, $P_{\rm rot}$ correlates with planetary mass loss rate. I therefore ran the models without a $P_{\rm rot}/\sin i$ input variable. The perturbed and combined success metrics are shown in Table 6.8. M1 and M2 explain 2 % and 4 % less variance respectively. A reduction in model performance is expected given the well-established relationship between activity and rotation period (Section 1.3.4). This may also indicate that the inclusion of planetary information improves model prediction success. The perturbed separation between the models increases to 2.93 σ. Faster planet mass loss rates increase

	ρ	R^2_a	RMSE	BIC	$\sigma_{res}[\%]$
Bootstrap					
M1					
μ_{BS1}	0.73	0.52	0.075	-378	65
σ_{BS1}	0.064	0.088	0.006	12.1	5.12
M2					
μ_{BS2}	0.66	0.41	0.084	-364	73
σ_{BS2}	0.070	0.093	0.006	11.1	5.21
S_{1-2}	0.68	0.88	-1.10	-0.83	-1.10
p_{1-2}	0.50	0.38	0.27	0.41	0.27
Perturbation method					
M1					
μ_{PM1}	0.71	0.49	0.078	-371	69
σ_{PM1}	0.024	0.036	0.003	5.3	2.40
M2					
μ_{PM2}	0.63	0.36	0.088	-352	78
σ_{PM2}	0.032	0.039	0.003	4.6	2.36
S_{1-2}	2.02	2.61	-2.63	-2.64	-2.63
p_{1-2}	0.044	0.009	0.008	0.008	0.008
Combined method					
M1					
μ_{C1}	0.70	0.50	0.077	-374	67
σ_{C1}	0.067	0.090	0.005	11.0	4.79
M2					
μ_{C2}	0.64	0.38	0.086	-362	75
σ_{C2}	0.069	0.089	0.006	11.2	5.44
S_{1-2}	1.96	2.38	-2.56	-1.97	-2.55
p_{1-2}	0.050	0.017	0.011	0.049	0.011
Permutation test					
μ_n	0.00	-0.68	0.143	-279	125
σ_n	0.115	0.186	0.008	8.7	7.00
M1					
$S_{1\text{-}n}$	5.56	6.06	-6.85	-6.67	-6.90
$p_{1\text{-}n}$	2.77e-08	1.34e-09	7.49e-12	2.59e-11	5.20e-12
M2					
$S_{2\text{-}n}$	4.94	5.43	-5.88	-6.10	-5.94
$p_{2\text{-}n}$	7.70e-07	5.55e-08	4.10e-09	1.06e-09	2.80e-09

TABLE 6.7: Success metrics (as described in the caption for Table 6.2) for OLS-kPCA model predictions. Metrics are also provided for the combined resampling approach.

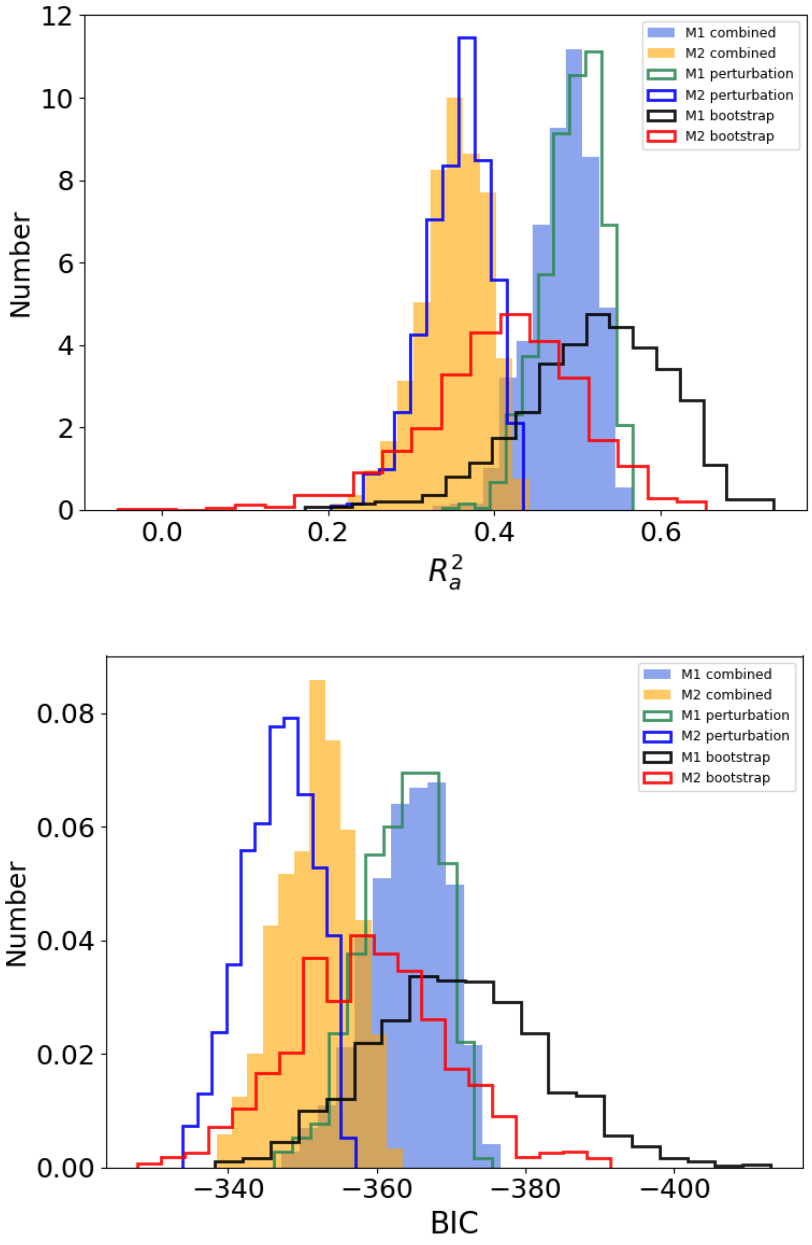

FIGURE 6.9: R_a^2 and RMSE distributions for OLS-kPCA Models 1 & 2. Each metric is presented in a separate panel. R_a^2 and larger negative BIC values indicate a stronger fit. Bootstrapped distributions are outlined in black and red, perturbed distributions are outlined green and blue, and combined distributions are filled blue and orange (see legends). Histogram counts are normalized.

enshrouding, which masks the effects of SPI. By removing $P_{\rm rot}$, M1 provides a purer model of SPI.

Outlier exclusion

The MS sample contains an outlier: the 36 $M_{\rm J}$ brown dwarf EPIC 219388192. When this target is removed from the input data, the mean perturbed $R^2_{\rm a}$ remains the same for M1 and M2. This indicates scaling and normalisation has been effective. The separation between bootstrapped distributions increases by 0.1 σ. Bootstrap resampling is expected to be more sensitive to outliers as some samples will not sample the outlier, while others will sample it multiple times, creating more spread.

	ρ	$R^2_{\rm a}$	RMSE	BIC	$\sigma_{\rm res}[\%]$
Perturbation method					
M1					
$\mu_{\rm PM1}$	0.70	0.47	0.081	-371	71
$\sigma_{\rm PM1}$	0.027	0.036	0.003	5.2	2.40
M2					
$\mu_{\rm PM2}$	0.60	0.32	0.091	-351	80
$\sigma_{\rm PM2}$	0.029	0.036	0.002	4.1	2.14
S_{1-2}	2.44	2.93	-2.93	-2.91	-2.93
p_{1-2}	0.015	0.003	0.003	0.004	0.003
Combined method					
M1					
$\mu_{\rm C1}$	0.69	0.47	0.079	-373	69
$\sigma_{\rm C1}$	0.061	0.077	0.005	10.5	4.71
M2					
$\mu_{\rm C2}$	0.60	0.34	0.089	-360	78
$\sigma_{\rm C2}$	0.069	0.083	0.006	10.8	5.43
S_{1-2}	2.25	2.52	-2.69	-2.07	-2.69
p_{1-2}	0.025	0.012	0.007	0.039	0.007

TABLE 6.8: Success metrics (as described in the caption for Table 6.2) for perturbed and combined distributions for OLS-kPCA models that do not include the $P_{\rm rot}/\sin i$ term.

6.5 Sub-sample tests

I ran the OLS-kPCA models on a series of sub-samples of interest, which included the two non-transiting targets where applicable. Perturbed distribution results are shown in Table 6.9. Apart from in the sub-basal population (where enshrouding is most likely to obscure true activity), M1 is invariably more successful. The OLS-kPCA M1 explains more than 50 % of the variance in the hot Jupiter (HJ: $M_{\rm P} > 0.157$ $M_{\rm J}$), close-in ($a < 0.05$ AU), strong interacting ($\log(M_P/a) > 0.5$), and large planetary orbital momentum ($j_{\rm orb} > j_{\rm rot}$) sub-samples. Separations of $\gtrsim 3\,\sigma$ are observed in the close-in and high activity ($\log(R_{\rm HK}) > -4.9$) populations.

For the HJ population, the models strengthen with $R^2_{\rm a} = 0.52$ and 0.42 for M1 and M2 respectively. With M2 strengthening, the S_{1-2} value drops and p rises, indicating $M_{\rm P}$ drives observed trends. For the closest-in population ($a < 0.05$ AU), there is a high $R^2_{\rm a}$ value of 0.57 for M1. Separation values vary between $2.99 - 3.59\,\sigma$ ($p = 2.94 \times 10^{-3} - 4.71 \times 10^{-4}$), showing a strong preference for inclusion of the planetary term. This provides evidence that SPI is at its most extreme and observable for the closest-orbiting systems. Success metrics remain solid for the 'extreme' populations ($\log(M_P/a) > 0.5$ and $M_P/a^2 > 450\,M_{\rm J}$ AU^{-2}), though $|S_{1-2}|$ values are more modest: between $0.88 - 1.15\,\sigma$. Where the planetary orbital angular momentum is greater than the stellar rotational momentum ($j_{\rm orb} > j_{\rm rot}$), 52 % and 45 % of variance is described respectively by M1 and M2, and $|S_{1-2}|$ approaches $1\,\sigma$.

For the low activity populations ($\log(R'_{\rm HK}) < -4.9$ and -5.1), the models are less successful. As seen in Chapter 5, slopes between planetary parameters and low activitiy measurements are generally flat. This may be because, with activity either genuinely dropping to basal levels or being obscured, variations are either minimal or masked. Despite being only 41-strong, the high activity population ($\log(R'_{\rm HK}) > -4.9$) has robust M1 success metrics (e.g. $R^2_{\rm a} = 0.39$) and $|S_{1-2}|$ approaching $4\,\sigma$ ($p = 8.52 \times 10^{-5}$). Enshrouding may be occurring to a lesser extend in this population, meaning true activity correlations are evident. M2 performs poorly for this sub-sample, explaining only 12 % of variance. To investigate this result I further optimized the OLS-kPCA model for this dataset, incorporating $P_{\rm rot}/\sin i$ and

distance d with 5 PC inputs for M1 and 4 PC inputs for M2. M1 improved to explain 48 % and M2 to account for 31 % of variance. $|S_{1-2}|$ values ranged from $2.11 - 2.16\,\sigma$. While this is a lower significance result, it is remarkable that a $2\,\sigma$ separation is apparent in a small sub-sample. Further augmenting the high activity sub-population may provide fascinating insight into SPI.

6.6 Predictor variable tests

The OLS-kPCA model was exploited to test whether a range of variables enhance or diminish prediction success. The results are shown in Table 6.10 for the MS and close-in ($a < 0.05$ AU) sub-samples. In the MS sample, there is $2.06 - 2.57\,\sigma$ separation between M1 and M2 models that incorporate $M_{\rm P}$, $M_{\rm P}/a$ and $h_{\rm tide}/h_{\rm scale}$ terms. Tidal SPI may be the main driver of activity variations. Separation is less significant for the more complex SPI term, $M_{\rm P}/a^2$, with $|S_{1-2}| = 1.03 - 1.28\,\sigma$. Similarly the magnetic SPI proxy ($1/a$) shows $|S_{1-2}| = 0.88 - 1.08\,\sigma$. Terms associated with mass loss ($T_{\rm eq}/g_P$, $g_{\rm P}$, $T_{\rm eq}$) have $R^2_{\rm a} = 0.39 - 42$, and $|S_{1-2}|$ values between $0.92 - 1.44\,\sigma$. Unsurprisingly mass loss signatures are less significant than SPI signals: higher mass loss rates lead to increased absorption and less observable activity variation.

In the close-in sub-sample, M1's prediction success improves, with $R^2_{\rm a} = 0.51 - 57$ when M_P, M_P/a, and $h_{\rm tide}/h_{\rm scale}$ are incorporated. This leads to $|S_{1-2}|$ values of 2.64, 3.23 and $3.97\,\sigma$ respectively, **providing strong evidence ($p = 0.008 - 3.27 \times 10^{-5}$) that M1 (incorporating these variables) and M2 do not explain the same variance.** SPI may undergo a step change in significance where $a < 0.05$ AU. Separation between the model distributions of $> 2\,\sigma$ is evident when $T_{\rm eq}$ is included (instead of the SPI term) in M1 for the close-in sub-sample. Although less significant than evidence for SPI, this result provides marginal proof of mass loss around high-activity hosts of close-orbiting planets. To further test the OLS-kPCA model on the high activity ($\log(R'_{\rm HK}) > -4.9$) sub-sample, I find that when M1 incorporates $M_{\rm P}$, $M_{\rm P}/a$ and $h_{\rm tide}/h_{\rm scale}$, it has superior prediction power to M2 by up to $\sim 4\,\sigma$. This confirms that, even in the less populous high activity sub-sample, including planetary variables significantly improves model performance.

TABLE 6.9: Sub-sample tests: success metrics for perturbed distribution results for OLS-kPCA M1 and M2 predictions for a variety of sub-samples. Numbers in brackets indicate the sample size.

	ρ	R^2_a	RMSE	BIC	σ_{res}
Main sequence (80)					
μ_1	0.71	0.48	0.080	-378	70
μ_2	0.63	0.35	0.089	-361	78
S_{1-2}	1.86	2.19	-2.20	-2.20	-2.21
p_{1-2}	0.063	0.029	0.028	0.028	0.028
Hot Jupiters (71)					
μ_1	0.74	0.52	0.081	-331	67
μ_2	0.67	0.42	0.089	-317	74
S_{1-2}	1.14	1.17	-1.18	-1.19	-1.18
p_{1-2}	0.253	0.244	0.238	0.233	0.238
$a < 0.05$ AU (54)					
μ_1	0.77	0.57	0.088	-242	64
μ_2	0.64	0.39	0.104	-224	76
S_{1-2}	2.99	3.43	-3.53	-3.59	-3.53
p_{1-2}	2.94e-3	8.68e-4	5.91e-4	4.71e-4	5.91e-4
$\log(M_P/a) > 0.5$ (70)					
μ_1	0.74	0.52	0.082	-324	67
μ_2	0.67	0.42	0.090	-311	74
S_{1-2}	1.14	1.13	-1.14	-1.15	-1.14
p_{1-2}	0.256	0.259	0.253	0.248	0.253
$M_P/a^2 > 450$ (40)					
μ_1	0.70	0.47	0.115	-150	71
μ_2	0.66	0.42	0.120	-147	74
S_{1-2}	1.14	0.88	-0.88	-0.89	-0.88
p_{1-2}	0.255	0.381	0.378	0.375	0.378
$j_{orb} > j_{rot}$ (31)					
μ_1	0.75	0.52	0.121	-117	67
μ_2	0.72	0.45	0.129	-113	72
S_{1-2}	0.69	0.99	-1.00	-1.00	-1.00
p_{1-2}	0.489	0.321	0.318	0.316	0.318
$\log(R'_{HK}) <$ -4.9 (39)					
μ_1	0.50	0.26	0.134	-142	84
μ_2	0.48	0.24	0.135	-141	85
S_{1-2}	0.20	0.22	-0.22	-0.22	-0.22
p_{1-2}	0.840	0.829	0.827	0.825	0.828

	ρ	R^2_a	RMSE	BIC	σ_{res}
$\log(R'_{HK})$ > -4.9 (41)					
μ_1	0.64	0.39	0.121	-150	76
μ_2	0.39	0.12	0.147	-135	92
S_{1-2}	3.44	3.93	-3.73	-3.52	-3.73
p_{1-2}	5.82e-4	8.52e-5	1.90e-4	4.35e-4	1.90e-4
$\log(R'_{HK})$ < -5.1 (24)					
μ_1	0.58	0.39	0.155	-77	76
μ_2	0.61	0.42	0.151	-78	74
S_{1-2}	-0.46	-0.39	0.40	0.41	0.40
p_{1-2}	0.642	0.694	0.689	0.683	0.689

TABLE 6.10: Success metrics for perturbed distributions where different planetary terms are incorporated into the OLS-kPCA M1 for MS and $a < 0.05$ AU sub-samples. M2 metrics detailed for the M_P term are the same for each variable.

	Main Sequence					**Close-in:** $a < 0.05$ **AU**				
	ρ	R^2_a	RMSE	BIC	σ_{res}	ρ	R^2_a	RMSE	BIC	σ_{res}
M_P										
μ_{PM1}	0.72	0.48	0.080	-377	70	0.74	0.51	0.093	-236	68
μ_{PM2}	0.62	0.35	0.089	-361	78	0.64	0.39	0.104	-224	76
S_{1-2}	2.66	2.57	-2.61	-2.64	-2.61	2.52	2.64	-2.76	-2.88	-2.76
p_{1-2}	0.008	0.010	0.009	0.008	0.009	0.012	0.008	0.006	0.004	0.006
M_P/a										
μ_{PM1}	0.71	0.48	0.080	-377	70	0.77	0.56	0.088	-241	64
S_{1-2}	1.84	2.16	-2.17	-2.18	-2.17	2.92	3.23	-3.33	-3.40	-3.33
p_{1-2}	0.066	0.031	0.030	0.029	0.030	0.003	0.001	0.001	0.001	0.001
M_P/a^2										
μ_{PM1}	0.67	0.43	0.083	-370	74	0.73	0.50	0.095	-234	69
S_{1-2}	1.03	1.27	-1.28	-1.28	-1.28	1.90	1.99	-2.04	-2.08	-2.04
p_{1-2}	0.302	0.205	0.202	0.199	0.202	0.057	0.047	0.042	0.037	0.042
$1/a$										
μ_{PM1}	0.66	0.41	0.085	-368	75	0.70	0.46	0.099	-230	72
S_{1-2}	0.88	1.07	-1.08	-1.08	-1.08	1.25	1.18	-1.20	-1.22	-1.20
p_{1-2}	0.378	0.284	0.279	0.281	0.281	0.210	0.238	0.230	0.223	0.230
h_{tide}/h_{scale}										
μ_{PM1}	0.69	0.45	0.082	-362	72	0.77	0.57	0.090	-231	64
S_{1-2}	1.78	2.06	-2.08	-2.09	-2.08	3.41	3.97	-4.06	-4.08	-4.06
p_{1-2}	0.075	0.039	0.038	0.037	0.038	3.98e-4	3.27e-5	2.25e-5	2.13e-5	2.25e-5
T_{eq}/g_P										
μ_{PM1}	0.67	0.40	0.087	-355	75	0.65	0.38	0.107	-212	76
S_{1-2}	1.19	1.14	-1.15	-1.15	-1.15	0.62	0.38	-0.39	-0.40	-0.39
p_{1-2}	0.235	0.255	0.252	0.249	0.252	0.538	0.701	0.696	0.690	0.696
g_P										
μ_{PM1}	0.65	0.39	0.087	-353	76	0.66	0.40	0.105	-214	75
S_{1-2}	0.93	0.92	-0.92	-0.93	-0.92	0.80	0.71	-0.72	-0.73	-0.72
p_{1-2}	0.352	0.359	0.356	0.353	0.356	0.425	0.481	0.474	0.467	0.474
T_{eq}										
μ_{PM1}	0.66	0.42	0.085	-357	75	0.71	0.48	0.099	-221	70
S_{1-2}	1.02	1.41	-1.42	-1.44	-1.42	2.05	2.28	-2.33	-2.36	-2.33
p_{1-2}	0.307	0.158	0.154	0.151	0.154	0.040	0.022	0.020	0.018	0.020

6.7 Summary

I have presented a study of star-planet interactions (SPI) and enshrouding in the OU-SALT MS sub-sample. I develop several predictive models to study these phenomena. Model validation and comparison statistics have been carefully chosen.

A multiple linear regression model based on France et al. (2018) was developed to predict stellar activity. Principle components analysis (PCA) was used to ensure predictor variables were independent. Ordinary least squares (OLS) regression was employed. Uncertainties on predictions were estimated with the perturbation technique. By exploiting the power of cross validation and several Monte Carlo-based methods, I have tested the statistical significance of model predictions.

Model 1 (M1), which included the simple SPI term $\log(M_{\mathbf{P}}/a)$**, accounted for** 36% **of variance in the data compared to** 28% **explained by Model 2 (M2), which did not contain a planetary term.** The mean Bayesian Information Criterion (BIC) score for M1 was higher by 9, and there was a $\sim 1.84\,\sigma$ ($p = 0.07$) separation in prediction success distributions, indicating a preference for the inclusion of the planetary term. Permutation tests showed the models were separated from the null distribution by $> 4-5\,\sigma$, providing very low probabilities ($10^{-5} - 10^{-7}$) that model prediction success is down to random chance.

The OLS-PCA model was improved by using Gaia DR2 data and optimizing input variables. I exploited non-linear/kernel PCA (kPCA) to make linear divisions in high dimensional feature space that—when projected back to 'normal' feature space—accounted for non-linearity in the data. I combined bootstrap resampling and the perturbation method to maximise the precision of my results. **The OLS-kPCA M1 accounted for** 49% **of variance compared to** 36% **for M2.** This equated to an improvement of 19 in the mean BIC score, and a 2.6 σ ($p = 0.009$) separation between perturbed model distributions. When $P_{\rm rot}/\sin i$ (which correlates with planetary mass loss) was removed from both models, **separation increased to** $> 2.93\,\sigma$ **(**$p = 0.003$**).**

The OLS-kPCA model was used to predict the activity of various sub-samples. Inclusion of the simple SPI term consistently improved prediction success. M1 explained 57 % of variance in the $a < 0.05$ AU sample, with R^2_a and BIC distributions separated by $\sim$ 3.5 σ from M2. Separations approached 4 σ in the high activity ($\log(R'_{HK}) > -4.9$) sub-sample, with a minimum p-value of 8.52×10^{-5}. The model was also used to test the inclusion of a range of planetary predictor variables. Separations between success metric distributions exceeded 2 σ when M_P or h_{tide}/h_{scale} terms were used in M1 instead of M_P/a. In the close-in sub-sample, separations exceed 4 σ when the h_{tide}/h_{scale} term was incorporated. These results indicate that it is highly unlikely ($p = 3.27 \times 10^{-5}$) that the models measure the same variance, and provide strong evidence that SPI influences stellar activity, particularly for the closest-orbiting systems.

The importance of accounting for interdependency of variables was noted in Chapter 5. To illustrate how this has been achieved, consider activity and stellar mass, which are highly correlated. There is also a well-defined relationship between stellar mass and planet mass. The $\log(R'_{HK}) - M_P$ correlation identified in Chapter 5 may have its origins in these correlations. My models account for both the $\log(R'_{HK}) - M_\star$ and $M_\star - M_P$ correlations as colour-index and effective temperature (both of which are correlated with $M_\star$) are included in M1 and M2. Any difference in performance of the models is due to the presence or absence of the planetary term.

I conclude that **in the MS and close-orbiting populations there is a** $0.03 - 10^{-5}$ **% chance that a model that incorporates** M_P**,** M_P/a **or** h_{tide}/h_{scale} **explains the same variance as a model that does not**. This indicates that the activity of close-in planet hosts is correlated with planetary variables. My results provide evidence for the action of SPI in these systems, and less significant proof of enshrouding. The close-in ($a < 0.05$ AU) and high activity ($\log(R'_{HK}) > -4.9$) samples provide particularly interesting test beds for these phenomena.

Collaboration details

I would like to thank Carole Haswell, John Barnes, Luca Fossati and Donal Hill for useful discussions about the analysis presented in this chapter.

Chapter 7

Conclusions

7.1 Thesis summary

7.1.1 Rotation period of HD 184960

HD 184960 is a bright, low activity, late-type F7V star (Spangler et al., 2001) identified as a close-in exoplanet host candidate by the Dispersed Matter Planet Project (DMPP: Haswell et al., 2019). To distinguish between planetary radial velocity signals and rotation-related jitter, it is desirable to determine the stellar rotation period.

We have acquired both B- and R-filter time-series photometric observations with the Las Cumbres Observatory Global Telescope (LCOGT). Defocusing was employed to avoid charge-coupled device (CCD) saturation and allow for longer exposure times. The Las Cumbres Observatory (LCO) pipeline reduces the data and performs source extraction. Quality checks were employed to search for systematic offsets between observations. None were identified. B-filter data was re-reduced using the Image Reduction and Analysis Facility (IRAF). Testing revealed optimal aperture radii of $8 - 10$ pixels. Photometry of R-filter data was carried out using AstroImageJ (AIJ) due to its image stacking functionality.

Three algorithms were employed to search for periodicity: the Lomb-Scargle (LS) periodogram; the string-length (SL) method; and phase dispersion minimization (PDM).

Periodic signals were identified by the LS algorithm in B-filter observations (Figure 2.5), with corresponding false alarm probabilities (FAPs) calculated using the bootstrap method. Notable periods included 5.1 d (FAP = 1.28×10^{-6}) and 6.1 d (FAP = 5.36×10^{-5}). The FAP values indicate that there is very low probability that signals occur by chance. However, overall LS power is low: $0.029 - 0.073$ for the most powerful periods. The signals are therefore treated with caution. The window function—which indicates periodicities that are due to data sampling alone—displayed a 1 d peak with power ≈ 0.8, as well as high power at longer periods. This indicated that intermediate period signals should be genuine. The SL method identified periods at 4.2 d ($p = 0.003$), 7.9 d ($p = 0.004$), 3.5 d ($p = 0.006$), 5.0 d ($p = 0.007$) and 7.0 d ($p = 0.007$). Many peaks were present, making it challenging to isolate the most promising candidate signal (Figure 2.7). PDM highlighted periods at 5.0 d ($p = 0.055$) and 6.3 d ($p = 0.014$) (Figure 2.8). Data was folded onto the most-promising periods, with $P_{\rm rot} = 5.1$ d producing the most sinusoidal-like binned, phase-folded light curve (Figure 2.9).

Similar analysis was undertaken for *R*-filter observations with periods of 5.3 d (FAP = 0.023) and 5.6 d ($p = 0.012$) emerging as the strongest candidate signals (Figure 2.14). Phase-folded light curves did not present as obviously sinusoidal as those of Pleiades stars (Hartman, 2010).

A range of possible rotation periods have thus been identified with no outstanding candidate emerging. Photometric amplitudes of Pleiades stars indicate why: for stars with $V-K \lesssim 1.1$, there is a step-drop in amplitude to $\lesssim 0.005$ mag (Rebull et al., 2016a). With $V-K = 1.2$, HD 184960 may be a low-amplitude star. I have constrained the amplitude of HD 184960 to less than 0.018 mag.

Identifying the rotation period of low activity planet hosts remains of key importance, particularly in radial velocity surveys. Using a multi-algorithmic approach to do so is encouraged.

7.1.2 The OU-SALT survey

Chromospheric activity measurements of planet hosts provide valuable information about the radiation environments that planets inhabit. However, exoplanet discovery papers often do not publish activity details. This is notably the case for planets detected with Kepler (Borucki et al., 2010), Super-WASP (Pollacco et al., 2006), and HATS-S (Bakos et al., 2004). When activity measurements are published, they are heterogeneous, having been measured by a variety of instruments and undergone different reduction processes. Activity catalogues, such as those published by Pace (2013), Figueira et al. (2014), and Boro Saikia et al. (2018) are beset by inconsistency, offsets and large calibration uncertainties.

The OU-SALT survey addresses these deficiencies. The survey generates activity measurements using the same telescope, spectrograph, instrument set-up, reduction process and calibration. **Measurement (internal) errors are on average three times smaller than uncertainties that include a calibration (external) error**. Studies with OU-SALT measurements may therefore be undertaken at higher resolution than prior state-of-the-art activity work.

Observations

The Robert Stobie Spectrograph (RSS) on the Southern African Large Telescope (SALT) was calibrated by Staab et al. (2017) to perform high precision measurements of Ca II H & K emission of close-orbiting, transiting planet hosts. Building on this work, I have generated activity measurements for 104 systems, 86 of which did not previously have published $\log(R'_{\rm HK})$ values (Figure 3.6). I selected bright ($M_{\rm V} < 13$) southern hemisphere targets with 4145 K < $T_{\rm eff}$ < 6800 K and $a < 0.11$ AU. The observed sample is comprised of systems that host 4 super Earths, 6 hot Neptunes, 93 hot Jupiters, and a brown dwarf. Eighteen targets in the OU-SALT sample have published $\log(R'_{\rm HK})$ values. A paired Student's t-test score of 0.19 indicates there is no systematic offset between OU-SALT measurements and published data.

System exploration

The activity distributions of main sequence (MS) and sub giant (SG) stars are distinct (Mittag et al., 2013; Staab et al., 2017). Classifying the evolutionary status of targets was a key exercise in this work. Gaia DR2 distances, calculated using a Bayesian framework by Bailer-Jones et al. (2018), were used to estimate absolute magnitudes. Targets were plotted on a Hertzsprung-Russell (HR) diagram with Gaia DR2 isochrones (Girardi et al., 2004, 2008; Dotter et al., 2008, Figure 4.4). The position of the terminal age of the main sequence (TAMS) extracted from Padova isochrones was also plotted. The eighty targets located below the TAMS were classified as MS, and the 24 above as SG.

Activity declines with age (Soderblom et al., 1991; Mamajek & Hillenbrand, 2008; Lorenzo-Oliveira et al., 2016). The ages of OU-SALT hosts were calculated using gyrochronolgy (Mamajek & Hillenbrand, 2008) and age-activity relations Barnes (2007). The majority of gyro ages are below 8 Gyr, indicating a relatively young population. In contrast, chromospheric ages stack up above 10 Gyr. Neither age calculation method is well-suited to close-orbiting systems (Figure 4.5). Combining the methods with isochrone dating may provide improved age estimates.

Star-planet interactions (SPI) are likely to be strong in OU-SALT systems, as indicated by several SPI proxies. The orbital angular momentum of the planet is more than the rotational angular momentum of the star ($J_{\rm orb} \gtrsim J_\star$) for around half of the OU-SALT sample (Figure 4.11). Approximately half of the systems also qualify as 'extreme' as defined by Miller et al. (2015), where the ratio of the planet mass to the semi-major axis squared exceeds 450 ($M_{\rm P}/a^2 > 450\ M_{\rm J}\ {\rm AU}^{-2}$, Figure 4.12). WASP-43 (Hellier et al., 2011), by way of example, has a $M_{\rm P}/a^2$ ratio more than 20 times this value. However, the systems are not so gravitationally distorted so as to approach filling their Roche lobes (Figure 4.10). WASP-103 (Gillon et al., 2012) has the highest Roche lobe filling factor of 54 %. OU-SALT planets are highly irradiated, receiving between 30 (K2-2: Vanderburg et al., 2015c) and 6400 (WASP-103) times the flux incident on Earth (Figure 4.8).

Thirteen targets have been observed multiple times. I measured the activity of ultra-hot Jupiter WASP-43 to be $\log(R'_{\rm HK}) = -4.18^{+0.09}_{-0.11}$. This confirms the measurement of Staab et al. (2017) ($\log(R'_{\rm HK}) = -4.17 \pm 0.10$). These anomalously high activity values likely result from star-planet tidal interactions (SPTI) rather than flares. Due to similarities with the WASP-12 system, WASP-103 was expected to display depressed activity. However, it is found to be a high-activity host with median activity $\log(R'_{\rm HK}) = -4.57 \pm 0.04$. This may be because the intense stellar wind generated by this hot star (with $T_{\rm eff} = 6110 \pm 100$ K) prevents the formation of a stable absorbing gas disk and/or inhibits mass loss (Vidotto & Cleary, 2020). KELT-11 (Pepper et al., 2017) displays an intriguing possible 100 d periodic signal (Figure 4.16). Further observations are required to confirm the significance of this period. Planet hosts HD 73256 (Udry et al., 2003), K2-32 (Sinukoff et al., 2016), (Hellier et al., 2012), WASP-67 (Hellier et al., 2012) and WASP-104 (Smith, A. M. S. et al., 2014) also display activity variations that warrant further follow-up observations.

Using a novel approach, I related orbital phase to activity for all survey observations—essentially treating all hosts as one high activity and one low activity target (Figure 4.19). Sinusoidal variation is possibly present in both high- and low-activity populations, indicating periodic star-planet interaction (SPI)-induced variation.

OU-SALT population study

To place the OU-SALT survey results into context, I compared them to two prominent activity catalogues: Pace (2013) (P13: Figure 5.3) and Boro Saikia et al. (2018) (BS18: Figure 5.2). On correcting a calibration error in BS18, I suggest that their conclusions concerning the diminished prominence of the Vaughan-Preston gap should be treated cautiously. The comparison with P13 provided the following key findings:

- Approximately a third of OU-SALT MS hosts display sub-basal activity, compared to only 2% of MS stars from P13;
- Depressed activity is also apparent in the OU-SALT SG sub-sample with $\log(R'_{\rm HK}) < -5.1$ for 75% of targets compared to 29% for P13 SG field stars; and

- A Hartigan's dip test p-value of 0.049 indicates bimodality in the OU-SALT MS sample.

These results provide evidence to prove our first two hypotheses: planets lose mass by atmospheric escape; and lost material forms diffuse clouds around star-planet systems that absorb emission flux. Bimodality in the MS sub-sample may reflect the presence of the Vaughan-Preston gap: high activity hosts efficiently strip mass and become enshrouded, while lower activity stars are less obscured.

Depressed activity values may result from absorption in the interstellar medium (ISM). To check whether this might be the case for OU-SALT hosts, I estimated extinction along the line of sight to each target using tools made available by Redfield & Linsky (2008) and Fossati et al. (2017b). I found $\log(R'_{\rm HK})$ may be reduced by up to 0.076, which does not account for the majority of sub-basal findings (Figure 5.6). Absorption by the ISM is unlikely to be the source of depressed activity values.

By identifying a sub-sample of systems where enshrouding and absorption is likely, I have been able to constrain the types of stars around which diffuse gas disks might form. Enshrouding occurs in systems with $T_{\rm eff} \approx 5200 - 6600$ K and $M_{\rm P} \approx 0.8 - 1.4\,M_{\odot}$ (Figures 5.18 and 5.19). The intense stellar wind of hotter stars blows away stripped material and/or inhibits mass loss. Cooler stars may provide insufficient thermal energy to prevent collapse of the disc, or generate insufficient extreme ultraviolet (EUV) flux for significant atmospheric escape.

I have undertaken an extensive and detailed search for correlations between planetary/stellar variables and activity in the OU-SALT sample. Combining internal (measurement) errors and the perturbation method provides uncertainties on correlation coefficients. The anti-correlation between $\log(R'_{\rm HK})$ and $g_{\rm P}^{-1}$ in the high activity ($\log(R'_{\rm HK}) > -4.9$) MS sub-sample ($\rho = -0.41 \pm 0.02$) is stronger than in the low activity ($\log(R'_{\rm HK}) < -4.9$) sub-sample by $\Delta\,\rho = 0.10$ (Figure 5.8). This is likely due to two factors that effect photoevaporation rates: planets orbiting active stars are exposed to higher EUV flux, leading to increased atmospheric escape; and lower gravity planets will more readily loose their atmospheres to space. This result confirms the findings of Fossati et al. (2015a), and provides further evidence to prove

our first two hypotheses.

Equilibrium temperature was calculated homogeneously for all targets but only weak $\log(R'_{\rm HK}) - T_{\rm eq}$ correlations were identified ($\rho = -0.07 \pm 0.01$: Figure 5.10). This may be because enshrouding and SPI effects both increase with $T_{\rm eq}$ but act in opposite senses on observed $\log(R'_{\rm HK})$.

Negative $\log(R'_{\rm HK}) - a$ correlations were identified with $\rho = -0.29 \pm 0.01$ for the MS sub-sample (Figure 5.11). For closer-in systems ($a < 0.05$ AU), $\rho = -0.43 \pm 0.01$ (MS) and $\rho = -0.33 \pm 0.02$ (SG). These correlations are stronger than reported by Martins et al. (2011) but weaker than those published by Fossati et al. (2015a). An anti-correlation between $\log(R'_{\rm HK})$ and a is present in the high-activity sub-sample ($\rho = -0.42 \pm 0.01$) but not in the low-activity sample ($\rho = 0.02 \pm 0.03$). Enshrouding may mask SPI-induced activity boosting in the low-activity sub-sample. Positive $\log(R'_{\rm HK}) - M_{\rm P}$ correlations are present, particularly for $M_{\rm P} > 0.8\,M_{\rm J}$, where $\rho = 0.52 \pm 0.01$ for the high-activity, MS sub-sample (Figure 5.12). The correlations are stronger than published by Canto Martins et al. (2011) and equivalent to those reported by Fossati et al. (2015b).

The commonly-cited SPI proxies ($\log(M_{\rm P})/a$ and $M_{\rm P}/a^2$) also exhibit significant correlations with $\log(R'_{\rm HK})$. For 'close-in, heavy planets', with $\log(M_{\rm P})/a > 0.6$ (Poppenhaeger et al., 2010), $\rho = 0.42 \pm 0.01$ (Figure 5.13). But for super-Earths and hot Neptunes ($\log(M_{\rm P})/a < 0.6$), a strong ($\rho = -0.53 \pm 0.03$) anti-correlation is present (albeit in a small sample). Strongly interacting systems have the highest activity. Similarly, a strong positive $\log(R'_{\rm HK}) - \log(M_{\rm P})/a^2$ correlation ($\rho = 0.34 \pm 0.01$) is present in the MS population, which strengthens to $\rho = 0.45 \pm 0.01$ for 'extreme' interacting systems with $M_{\rm P}/a^2 > 450\,M_{\rm J}\,{\rm AU}^{-2}$ (Miller et al., 2015) (Figure 5.14).

These correlations provide *prima facie* evidence of SPI in the OU-SALT systems. However, a number of studies highlight the importance of accounting for the interdependency of variables and selection effects (Poppenhaeger et al., 2010; Miller et al., 2015; France et al., 2018). Indeed, the $\log(R'_{\rm HK}) - a$ ($\rho = -0.29 \pm 0.01$) correlation identified may have its origin in $M_\star - a$ ($\rho = 0.39 \pm 0.01$) and $\log(R'_{\rm HK}) - M_\star$ ($\rho = -0.37 \pm 0.02$) correlations.

Untangling an interconnected web of variables to identify a genuine relationship between activity and planetary parameters was the focus of Chapter 6.

Multivariate analysis

I undertook a detailed multivariate analysis to search for signals of SPI and enshrouding in the OU-SALT MS sample. To begin, I built a multiple linear regression model to predict host activity. As per France et al. (2018)—who searched for SPI signals in far-ultraviolet (FUV) activity measurements—the following predictor variables were chosen: maximum rotation period ($P_{\rm rot}/\sin i$), absolute magnitude ($M_{\rm V}$), colour index ($B-V$), distance between the observer and star (d), effective temperature ($T_{\rm eff}$), and a simple SPI term ($M_{\rm P}/a$, where $M_{\rm P}$ is planet mass and a is separation).

Multiple linear regression requires uncorrelated predictor variables. As some of the chosen input variables are strongly correlated (e.g. $B-V$ and $T_{\rm eff}$), principle components analysis (PCA) was used to generate independent input variables (principal components) and filter out less useful information. Weighting coefficients of the multiple linear regression model were calculated using ordinary least squares (OLS) regression. As such, I called this the OLS-PCA model.

Two versions the OLS-PCA model were run in parallel: Model (M1) used the variables described above, while Model 2 (M2) dropped the SPI term. If M1 makes better predictions, then the planet influences chromospheric activity. If there is no improvement in prediction success, or indeed if M2 makes better predictions, then the planet has no significant influence on activity.

The following metrics were used to assess model success: Spearman's rank coefficient (ρ), root-mean-square error (RMSE), adjusted R^2 ($R^2_{\rm a}$), the Bayesian Information Criterion (BIC), and the standard deviation of the distribution of residuals between predicted and observed activity calculated as a percentage of the standard deviation of measured activity ($\sigma_{\rm res}$). I also employed three Monte Carlo resampling methods

to assess model performance: bootstrap, the perturbation technique, and permutation testing. Success metrics were calculated for each sample, allowing the construction of distributions.

M1 and M2 accounted for 36 % and 28 % of variance respectively (Table 6.2). M1's mean BIC score was higher by 9. M1 and M2 success metric distributions were separated by up to $1.84\,\sigma$ ($p = 0.07$). These results indicate that M1 (which incorporates a planetary parameter) performs more favourably than M2, with a 7 % probability that the difference between the models is not significant (Figure 6.5). The separation between the null distribution (as calculated with a permutation test) and model distributions reached $5.35\,\sigma$. This means there is a very low probability ($p = 8.72 \times 10^{-6}$) that the OLS-PCA models are describing noise. The null hypothesis may be rejected.

Several improvements to the OLS-PCA model were made: M_V, $B-V$ and $T_{\rm eff}$ were replaced by Gaia DR2 values for colour index ($R_G - B_G$), absolute magnitude (G_V), and effective temperature ($T_{\rm eff\,G}$). This provided more homogeneous inputs and avoided systematic offsets for these input parameters. Kernel PCA (with a cosine kernel) was used instead of PCA to account for non-linearities in the data. Distance d was dropped from the analysis as its inclusion reduced the models' ability to explain variance by 1-2 %. I called this improved version the OLS-kPCA model.

The OLS-kPCA M1 and M2 explained 49 % and 36 % of variance respectively (Table 6.7). M1's BIC score was higher by 19. Success metric distributions were separated by up to $2.64\,\sigma$, indicating a 0.8 % chance that M1 and M2 explain the same variance (Figure 6.9). When $P_{\rm rot}/\sin i$ (which correlates with mass loss rate) was dropped from the models, separations increased to $2.93\,\sigma$ ($p = 0.003$) (Table 6.8). The model that incorporates a planetary parameter better describes host activity than the model that does not, providing robust evidence that the planet influences activity.

Larger differences between M1 and M2 performance are evident for several subsamples. For example, in the $a < 0.05$ AU sub-sample, separations between M1 and M2 distributions reach $3.59\,\sigma$ ($p = 4.71 \times 10^{-4}$) (Table 6.9). Intense SPI interactions at close separations may provide strong activity variations. In the high-activity ($\log(R'_{\rm HK}) > -4.9$) sub-sample, distributions are separated by up to $3.93\,\sigma$

($p = 8.23 \times 10^{-5}$). While higher mass loss rates are expected in this sub-sample, enshrouding may be occurring to a lesser extent, meaning true activity variations are observable.

I also input a range of different variables into M1 in the place of the simple SPI term. Substituting $M_{\rm P}$ and $h_{\rm tide}/h_{\rm scale}$ also produced separations that exceeded $> 2\,\sigma$ (Table 6.10). Indeed, separation exceeds $4\,\sigma$ ($p = 2.13 \times 10^{-5}$) where $h_{\rm tide}/h_{\rm scale}$ is used in M1 for the close-in sub-sample. This provides evidence that tidal SPI influences host activity and may be the main driver of the correlations identified. I have therefore produced evidence of $2 - 4\,\sigma$ ($p \approx 0.003 - 10^{-5}$) significance that close-in orbiting exoplanets impact host activity. This provides proof of our third hypothesis.

I have been able to identify SPI where other studies (e.g. Kashyap et al., 2008; Martins et al., 2011; Miller et al., 2015; Fossati et al., 2015b; France et al., 2018) have not for several reasons. The homogeneous nature of the OU-SALT survey—with small internal uncertainties—permits a high resolution search. We have targeted the most extreme systems, which are the most likely to experience mass loss and SPI. The transit method is less sensitive to host activity than the radial velocity method, meaning there is less bias towards low activity hosts within the study. Using PCA allows independent predictor variables to be input into the multivariate analysis. Adopting non-linear kPCA has allowed my work to account for non-linearities. Using a two-model approach addresses the issue of interdependency of variables and selection bias. Finally, I adopt a considerably more rigorous approach to uncertainty analysis than previous studies (e.g. France et al., 2018). I use a range of success metrics and embrace frequentist methodologies to quantify the significance of my findings.

7.2 Limitations

Rotation periods

Several factors precluded a high-confidence identification of HD 184960's rotation period. The use of multiple instruments and limited exposure times did not provide adequate observing stability. The timespan of observations could also have played a role: small spots that induce modest photometric changes may have evolved over the time scale of observations, confounding identification of the rotation period.

OU-SALT survey

The statistical power of the analysis presented is limited by sample size. OU-SALT observations are ongoing and the sample continues to grow. Target selection criteria have mainly been applied arbitrarily. The study to-date has given limited resource to time-series observations of individual targets. Observing hosts with highly variable activity over multiple epochs will improve understanding of activity variation. For the lowest activity targets, emission flux in Ca II H & K cores approaches zero, inevitably leading to lower signal-to-noise ratio (SNR) and larger uncertainties. Any incremental improvements in flux extraction may improve SNR.

As encouraged by Luri et al. (2018), Bayesian-estimated distances (from BJ18) were used to calculate absolute magnitudes. BJ18 do however emphasise that better results could be achieved for individual objects with a tailored approach. Given luminosity $L \propto d^2$, errors in d could be sufficient to change MS/SG classification in borderline cases. Ideally, bespoke Bayesian analysis should be undertaken to calculate the distance to each target. In addition, I used metallicity categories when classifying targets. Plotting metallicity specific isochrones for each target would likely improve classification. Similarly, it is preferable to undertake ISM absorption analysis for each individual target, such as that undertaken for WASP-13 by Fossati et al. (2015b). This would be resource expensive but provide a preferable approach that could be adopted for targets of particular interest. Combining gyrochronology and isochrone dating may improve age estimates.

Testing a wider range of regression algorithms may improve performance of the multivariate model. For example, support vector machine regression is known to be effective for smaller data sets (Vapnik & Chervonenkis, 1964). BS18 and/or P13 data could be used to train algorithms to predict non-host activity and distinguish the activity of systems with/without planets. To this end, it would be useful to know which BS18 systems are close-in planet hosts. Other quantitative structures, such as variational Bayesian inference (Roberts et al., 2013), may provide powerful insights. It might also be illuminating to search for causal mechanisms in systems, using causal inference (Irzik & Meyer, 1987).

7.3 Future work

Rotation periods of low mass planet hosts

The DMPP will continue to discover low mass planets orbiting low activity stars using the radial velocity technique. Identifying stellar rotation periods remains crucial to ruling out alias signals. Pinpointing the rotation period of HD 184960 was particularly difficult due to its brightness and low photometric amplitude. High cadence observations spanning several rotation periods with a stable instrument, such as the Transiting Exoplanet Survey Satellite (TESS: Ricker et al., 2014), will help identify the rotation period of low amplitude targets. The multi-algorithmic approach taken to identifying rotation periods is robust. However, other algorithms—such as the autocorrelation function used by McQuillan et al. (2013a)—could also be trialled.

Extending the OU-SALT spectroscopic survey

There is scope to extend the OU-SALT survey in several ways. Targeting systems with observed evidence of mass loss and estimated mass loss rates (such as the systems with detection of the helium triplet details in Section 1.4.1) will permit the relationship between mass loss and activity to be better quantified.

Newly discovered planet hosts that fall within our selection criteria (Section 4.1) will provide exciting new targets. For example, the Next Generation Transit Survey

(NGTS: Wheatley et al., 2018) has already detected some fascinating targets, such as NGTS-10—a HJ in a 0.8 d orbit around a K5V star (McCormac et al., 2020). Low-mass transiting planets revealed by TESS will provide a rich source of targets for this survey.

Our selection criteria could also be relaxed to investigate hosts of further-out planets. This would allow us to probe how mass loss and SPI tail-off with separation and $T_{\rm eq}$. The survey might also be extended to hosts of non-transiting planets. This will help investigate whether diffuse gas disks are confined to the plane of orbit, or whether they enshroud the entire system. It would also allow us to better investigate selection effects.

We will perform follow-up studies on interesting systems revealed by our survey. For example, considerable $\log(R'_{\rm HK})$ variation is expected for high activity host WASP-43 but our two observations to date are remarkably similar. It would be interesting to observe activity variations of this target over time. Further observations are required to confirm the possible $\sim$ 100 d activity cycle of KELT-11. WASP-103, HD 73256, K2-32, WASP-63, WASP-67 and WASP-104 all displayed variation to justify follow-up. In general, high activity targets will show more variation and may be less obscured by enshrouding gas clouds, so should be prioritised for SPI studies.

Activity in M-dwarfs: the calcium triplet

The prominent lines found at 8498, 8542 and 8662 Å in the near infrared region of G, K and M star spectra are known as the Ca II infrared triplet (IRT). They are formed in the lower chromosphere and may be used as indicators of chromospheric activity (Linsky et al., 1979; Montes et al., 2000; Chmielewski, 2000). Radiative losses of Ca II H & K and IRT lines are expected to be related as they share the same upper excited state (Linsky et al., 1979). Indeed, Martin et al. (2002) find the excess flux in these line cores to be well correlated, provided $B-V$ dependency is accounted for. Andretta et al. (2005) define the $R_{\rm IRT}$ index, which is the ratio of the central line depression to a photospheric contribution, and the analogue of $R'_{\rm HK}$.

Flux distributions of M stars peak at longer wavelengths, meaning the Ca II IRT provides a window through which the chromospheric activity of the most common type of star may be monitored. The GAIA mission (Prusti, 2011) and CARMENES (Quirrenbach et al., 2014) spectrograph have wavelength coverage of the Ca II IRT lines but not H & K lines. Our activity survey could be extended to cover planet-hosting dwarfs which, as noted in the introduction, have been observed to host some fascinating planetary systems (e.g. TRAPPIST-1: Gillon et al., 2017).

Transmission spectroscopy

OU-SALT and DMPP systems are prime candidates for follow-up observations that seek to characterise planets. Transmission spectroscopy could be used to search for chemical signatures in stellar light that has passed through diffuse gas clouds of material stripped from planets. Indeed, this has been undertaken for some of the ultra hot Jupiters in the OU-SALT survey, such as WASP-43 (Murgas et al., 2014) and WASP-103 (Lendl et al., 2017), with fascinating results. While seeking to identify atmospheric make-up, this could also present the opportunity to characterise the bulk composition of exoplanets (Rappaport et al., 2012; Perez-Becker & Chiang, 2013). Instruments such as the Atmospheric Remote-sensing Infrared Exoplanet Large-survey (ARIEL: Tinetti et al., 2016) make such follow-up studies possible.

Appendix A

Principal components analysis & model success

A.1 Linear PCA

This derivation follows Section 12.1 of Bishop (2015), who states that "the aim of PCA is to project data onto a space with dimensionality $M < W$ while maximizing the variance". This involves finding M eigenvectors, being the principal components, corresponding to M largest eigenvalues of the correlation matrix $\mathbf{Z}$. Taking data that is projected onto one-dimensional space ($M = 1$), with direction defined by unit vector $\mathbf{u}_1$, each point is projected onto a scalar value $\mathbf{u}_1^T \mathbf{x}_n$.

The mean of the data is $\mathbf{u}_1^T \overline{\mathbf{x}}$ where $\overline{\mathbf{x}}$ is the sample mean given by:

$$\overline{\mathbf{x}} = \frac{1}{N} \sum_{n=1}^{N} \mathbf{x}_n; \tag{A.1}$$

and the variance of the projected data is:

$$\frac{1}{N} \sum_{n=1}^{N} \left\{ \mathbf{u}_1^{\mathrm{T}} \mathbf{x}_n - \mathbf{u}_1^{\mathrm{T}} \overline{\mathbf{x}} \right\}^2 = \mathbf{u}_1^{\mathrm{T}} \mathbf{Z} \mathbf{u}_1; \tag{A.2}$$

where $\mathbf{Z}$ is the data covariance/correlation matrix defined by:

$$\mathbf{Z} = \frac{1}{N} \sum_{n=1}^{N} \left(\mathbf{x}_n - \overline{\mathbf{x}} \right) \left(\mathbf{x}_n - \overline{\mathbf{x}} \right)^{\mathrm{T}}. \tag{A.3}$$

To maximize the projected variance of $\mathbf{u}_1^{\mathrm{T}}\mathbf{Z}\mathbf{u}_1$ with respect to $\mathbf{u}_1$, a Lagrange multiplier λ_1 enforces the normalization condition $\mathbf{u}_1^T\mathbf{u}_1 = 1$ using an expression for unconstrained maximization:

$$\mathbf{u}_1^{\mathrm{T}}\mathbf{Z}\mathbf{u}_1 + \lambda_1\left(1 - \mathbf{u}_1^{\mathrm{T}}\mathbf{u}_1\right). \tag{A.4}$$

Setting the derivative with respect to $\mathbf{u}_1 = 0$, the quantity has a stationary point when:

$$\mathbf{Z}\mathbf{u}_1 = \lambda_1\mathbf{u}_1. \tag{A.5}$$

This means $\mathbf{u}_1$ is an eigenvector of $\mathbf{Z}$. Multiplying by $\mathbf{u}_1^T$ and noting that $\mathbf{u}_1$ is a unit vector (so $\mathbf{u}_1^T\mathbf{u}_1 = 1$), the variance is given by:

$$\mathbf{u}_1^{\mathrm{T}}\mathbf{Z}\mathbf{u}_1 = \lambda_1. \tag{A.6}$$

The variance is maximum when $\mathbf{u}_1$ is equal to the eigenvector with the largest eigenvalue λ_1. This eigenvector is the first principal component. In the generalized case, PCs are defined by the eigenvectors $\mathbf{u}_i$ of the covariance/correlation matrix:

$$\mathbf{Z}\mathbf{u}_i = \lambda_i\mathbf{u}_i, \tag{A.7}$$

where $i = 1, ..., W$, and the $W \times W$ sample correlation matrix $\mathbf{Z}$ is defined by:

$$\mathbf{Z} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}_n\mathbf{x}_n^{\mathrm{T}}. \tag{A.8}$$

Eigenvectors are normalized such that $\mathbf{u}_i^T\mathbf{u}_i = 1$. Maximum variance is defined by M eigenvectors $\mathbf{u}_1, ..., \mathbf{u}_M$ of the data covariance/correlation matrix $\mathbf{Z}$, corresponding to M largest eigenvalues $\lambda_1, ..., \lambda_M$.

A.2 Kernel PCA

This derivation follows Section 12.3 of (Bishop, 2015). The aim of non-linear PCA is to transform each data point onto $\mathbf{x}_n \rightarrow \phi\mathbf{x}_n$ in M-dimensional feature space. Data vectors may be expressed in the form of scalar products $\mathbf{x}_n^T\mathbf{x}_m$. Assuming the projected data has zero mean $\sum_n \phi(\mathbf{x}_n) = 0$, the $M \times M$ sample correlation matrix is given by:

$$\mathbf{C} = \frac{1}{N}\sum_{n=1}^{N} \phi\left(\mathbf{x}_n\right)\phi\left(\mathbf{x}_n\right)^{\mathrm{T}}; \tag{A.9}$$

and its eigenvector expansion defined by:

$$\mathbf{C}\mathbf{v}_i = \lambda_i \mathbf{v}_i, \tag{A.10}$$

where $i = 1, ..., M$. Eigenvalues may be calculated without working in the original feature space. Equation A.10 shows:

$$\frac{1}{N}\sum_{n=1}^{N} \phi\left(\mathbf{x}_n\right)\left\{\phi\left(\mathbf{x}_n\right)^{\mathrm{T}}\mathbf{v}_i\right\} = \lambda_i \mathbf{v}_i; \tag{A.11}$$

so the vector $\mathbf{v}_i$ is given by a linear combination of of $\phi(\mathbf{x}_n)$ and so can be written:

$$\mathbf{v}_i = \sum_{n=1}^{N} a_{in}\phi\left(\mathbf{x}_n\right). \tag{A.12}$$

Substituting this back into the eigenvector equation:

$$\frac{1}{N}\sum_{n=1}^{N} \phi\left(\mathbf{x}_n\right)\phi\left(\mathbf{x}_n\right)^{\mathrm{T}}\sum_{m=1}^{N} a_{im}\phi\left(\mathbf{x}_m\right) = \lambda_i \sum_{n=1}^{N} a_{in}\phi\left(\mathbf{x}_n\right). \tag{A.13}$$

Multiplying both sides by $\phi(\mathbf{x}_l)^T$:

$$\frac{1}{N}\sum_{n=1}^{N} k\left(\mathbf{x}_l\mathbf{x}_n\right)\sum_{m=1}^{m} a_{im}k\left(\mathbf{x}_n\mathbf{x}_m\right) = \lambda_i \sum_{n=1}^{N} a_{in}k\left(\mathbf{x}_l\mathbf{x}_n\right). \tag{A.14}$$

expresses the eigenvector equation in terms of the kernel function. This may, in turn, be expressed:

$$\mathbf{K}^2\mathbf{a}_i = \lambda_i N\mathbf{K}\mathbf{a}_i, \tag{A.15}$$

where $\mathbf{a}_i$ is an N-dimensional column vector with elements a_{ni} for $n = 1, ..., N$. To find solutions for $\mathbf{a}_i$, we divide through by $\mathbf{K}$, then solve:

$$\mathbf{K}\mathbf{a}_i = \lambda_i N\mathbf{a}_i. \tag{A.16}$$

Using Equations A.12 and A.16 to normalize the eigenvectors in feature space:

$$1 = \mathbf{v}_i^{\mathrm{T}}\mathbf{v}_i = \sum_{n=1}^{N}\sum_{m=1}^{N} a_{in}a_{im}\phi\left(\mathbf{x}_n\right)^{\mathrm{T}}\boldsymbol{\phi}\left(\mathbf{x}_m\right) = \mathbf{a}_i^{\mathrm{T}}\mathbf{K}\mathbf{a}_i = \lambda_i N\mathbf{a}_i^{\mathrm{T}}\mathbf{a}_i. \tag{A.17}$$

Then $\mathbf{x}$ is projected onto the eigenvector i using Equation A.12:

$$y_i(\mathbf{x}) = \phi(\mathbf{x})^{\mathrm{T}}\mathbf{v}_i = \sum_{n=1}^{N} a_{in}\phi(\mathbf{x})^{\mathrm{T}}\phi\left(\mathbf{x}_n\right) = \sum_{n=1}^{N} a_{in}k\left(\mathbf{x}\mathbf{x}_n\right). \tag{A.18}$$

After centralizing (assuming a non-zero mean), the projected data points are given by:

$$\widetilde{\phi}\left(\mathbf{x}_n\right) = \phi\left(\mathbf{x}_n\right) - \frac{1}{N}\sum_{l=1}^{N}\phi\left(\mathbf{x}_l\right). \tag{A.19}$$

The corresponding Gram matrix (being the Hermitian matrix of all possible inner products) is:

$$\begin{aligned}
\widetilde{K}_{nm} =& \widetilde{\phi}\left(\mathbf{x}_n\right)^{\mathrm{T}} \widetilde{\phi}\left(\mathbf{x}_m\right) \\
=& \phi\left(\mathbf{x}_n\right)^{\mathrm{T}} \phi\left(\mathbf{x}_m\right) - \frac{1}{N}\sum_{l=1}^{N} \phi\left(\mathbf{x}_n\right)^{\mathrm{T}} \phi\left(\mathbf{x}_l\right) \\
& - \frac{1}{N}\sum_{l=1}^{N} \phi\left(\mathbf{x}_l\right)^{\mathrm{T}} \phi\left(\mathbf{x}_m\right) + \frac{1}{N^2}\sum_{j=1}^{N}\sum_{l=1}^{N} \phi\left(\mathbf{x}_j\right)^{\mathrm{T}} \phi\left(\mathbf{x}_l\right). \\
=& k\left(\mathbf{x}_n \mathbf{x}_m\right) - \frac{1}{N}\sum_{l=1}^{N} k\left(\mathbf{x}_l \mathbf{x}_m\right) \\
& - \frac{1}{N}\sum_{l=1}^{N} k\left(\mathbf{x}_n \mathbf{x}_l\right) + \frac{1}{N^2}\sum_{j=1}^{N}\sum_{l=1}^{N} k\left(\mathbf{x}_j \mathbf{x}_l\right)
\end{aligned} \tag{A.20}$$

In matrix notation:

$$\tilde{\mathbf{K}} = \mathbf{K} - \mathbf{1}_N \mathbf{K} - \mathbf{K}\mathbf{1}_N + \mathbf{1}_N \mathbf{K} \mathbf{1}_N, \tag{A.21}$$

where $\mathbf{1}_N$ is the $N \times N$ matrix in which every element takes the value 1/N. $\tilde{\mathbf{K}}$ may be used to find eigenvalues and eigenvectors. Using a linear kernel returns the linear PCA result.

A.3 Types of kernel

There is a wide variety of kernel functions, the simplest example of which is the linear kernel:

$$k(\mathbf{x}, \mathbf{x}') = \mathbf{x}^T \mathbf{x}'. \tag{A.22}$$

The Gaussian radial basis function (RBF) is, perhaps, the most widely used kernel:

$$k(\mathbf{x}, \mathbf{x}') = \exp(-\gamma ||\mathbf{x} - \mathbf{x}'||^2), \tag{A.23}$$

where $||\mathbf{x} - \mathbf{x}'||^2$ is the squared Euclidean distance between two feature vectors and $\gamma = 1/2\, b^2$, where b^2 is known as the bandwidth parameter, and is to be optimized.

The cosine kernel function can also play an important role in separating feature vectors. The cosine between two vectors is the ratio of the dot product to the product of the individual norms:

$$k(\mathbf{x}, \mathbf{x}') = \cos\theta = \left\langle \frac{\mathbf{x}}{||\mathbf{x}||}, \frac{\mathbf{x}'}{||\mathbf{x}'||} \right\rangle. \tag{A.24}$$

The sigmoid kernel between two vectors is defined:

$$k(\mathbf{x}, \mathbf{x}') = \tanh(\gamma \mathbf{x}\mathbf{x}' + C_0), \tag{A.25}$$

where γ is the slope and C_0 is the y-intercept. The polynomial kernel is defined:

$$k(\mathbf{x}, \mathbf{x}') = \tanh(\gamma \mathbf{x}\mathbf{x}' + C_0)^d, \tag{A.26}$$

where d is the kernel degree. Polynomial kernels consider the similarity between vectors across dimensions.

A.4 Assessing model success

The aim of model fitting and parameter estimation is to provide the "most parsimonious best fit of a mathematical model to data" (Feigelson & Babu, 2012). A preliminary test involves assessing whether variables are correlated using coefficients, such as the Spearman's rank coefficient or Kendall's tau (discussed in Section 5.4).

Three approaches emerge for validating and comparing models:

1. Examine plots of observation-prediction residuals;
2. Calculate global measures of model success including residual-based statistics and likelihood-based information criteria; and
3. Undertake cross validation which, being particularly relevant to machine-learning algorithms, is discussed in Section B.2.3.

A.4.1 Residual-based analysis

Residuals between measured and predicted activity ($\hat{e}_n = y_n - \hat{y}_n$) may be plotted to check for discrepancies and outliers (Kutner, 2005; Sheather, 2009; Feigelson & Babu, 2012). A successful model should account for a large fraction of scatter. The root-mean-square error (RMSE) measures the difference between predicted and observed values. It is the standard deviation of the residuals:

$$\text{RMSE} = \sqrt{\sum_{n=1}^{N} \frac{(\hat{y}_n - y_n)^2}{N}}. \tag{A.27}$$

Another commonly used residual-based statistic is the coefficient of determination (R^2), which details the ratio of the error sum of squares to the total sum of squares:

$$R^2 = 1 - \frac{\sum_{n=1}^{N} (y_n - \hat{y}_n)^2}{\sum_{n=1}^{N} (y_n - \overline{y}_n)^2}, \tag{A.28}$$

where $\overline{y}_n = \Sigma y_n / N$ is the mean of the dependent variable. While adding parameters (i.e. increasing W) may improve the correlation indicated by R^2, it may also lead to

over-fitting. Model selection by maximizing R^2 is thus ill-advised. Instead, adjusted R^2 (R_a^2: Theil, 1961), which penalizes the number of parameters, may be used:

$$R_a^2 = 1 - \frac{N-1}{N-W} R^2. \tag{A.29}$$

A.4.2 Likelihood-based metrics

Two prominent likelihood-based estimators of the quality of statistical models for a *given set of data* are the Akaike Information Criterion (AIC: Akaike, 1974), and the Bayesian Information Criterion (BIC: Schwarz, 1978). These metrics, which estimate the quality of a model relative to other models, are calculated:

$$\begin{aligned} \text{AIC} &= 2W - \text{In}(\hat{L}); \ \& \\ \text{BIC} &= \text{In}(N)W - 2\text{In}(\hat{L}), \end{aligned} \tag{A.30}$$

where the $-\text{In}(\hat{L})$ terms are the log-likelihood functions that describe the probability of predicting the observed data, and the W terms penalise the model for the addition of new parameters. Note the difference between the strength of the penalty for increasing model parameters: the BIC penalty scales with both the number of parameters and sample size, imposing a harsher penalty. The magnitude of the difference of BIC values illustrates to what extent one model is preferred over another: $\Delta\,\text{BIC} = 0 - 2$ provides weak evidence of model improvement; $\Delta\,\text{BIC} = 2 - 6$ provides positive evidence; $\Delta\,\text{BIC} = 6 - 10$ provides strong evidence; and $\Delta\,\text{BIC} > 10$ provides very strong evidence (Raftery, 1995).

Appendix B

Supervised learning approach

B.1 Introduction

To attempt to improve model prediction success, I used a supervised learning approach. Machine learning (ML) algorithms map predictor variables to a target variable in order to train a model that makes predictions about unseen data. While a multitude of supervised learning algorithms are available (e.g. Linear Discriminant Analysis, Naïve Bayes, SVMs, random forests and neural networks), boosted decision tree algorithms have recently proven to be extremely powerful prediction tools.

Decision tree algorithms generate predictions by making a series of binary decisions within a tree-like structure. Trees consist of a network of nodes, each of which has two branches (apart from the 'leaf' node—the final node along a decision path: Figure B.1). A regression tree assigns a number to a pattern of predictor variables by filtering inputs down through separate paths in the tree. Tree-based models are "readily interpretable by humans because they correspond to a sequence of binary decisions applied to individual predictor variables"—Bishop (2015).

Tree *boosting* is an ensemble technique where multiple learning algorithms are exploited to obtain better predictive performance (Friedman, 2001). New models are added to correct errors made by existing models until no further improvements can be made. Thus 'weak learners' are iteratively combined into a single 'strong learner'.

Boosting was extended to regression by (Friedman, 2001). In gradient boosting, new models are added that predict the residuals or errors of prior models, summing them

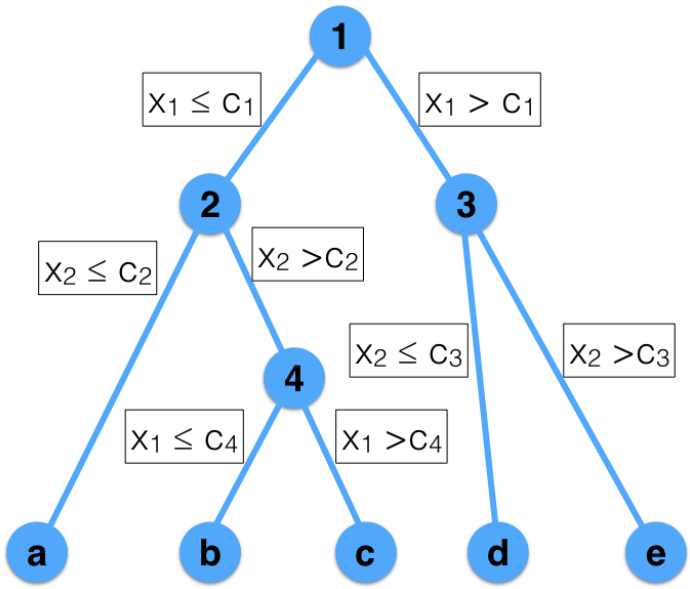

FIGURE B.1: Diagrammatic depiction of a binary decision tree. The input space is divided in two at Node 1 according to whether $x_1 \leq c_1$ or $x_1 > c_1$, where c_1 is a model parameter. This creates two sub-areas, each of which can be independently divided. For example, at Node 2 the sub-space is divided according to whether x_2 is less or more than c_2. The destination of a new input x is determined by the path it takes from the root node to a specific leaf node ($a - e$), according to decision criteria at each node. (Adapted from Bishop, 2015)

to make a final prediction. A first-order iterative optimization algorithm, known as a gradient descent algorithm, is used to minimize a differential loss function when adding new models.

There are several commonly used gradient-boosting decision tree algorithms. Adaptive Boosting (Adaboost: Freund & Schapire, 1997) adopts a sequential ensemble technique where n learners are produced while the training set is selected by random sampling with replacement. A more recently developed algorithm, eXtreme Gradient Boosting or *XGBoost* (Chen & Guestrin, 2016), is noted for its execution speed and performance, and has proven extremely successful in analysing structured data, notably in competitions on the Kaggle data science platform[1]. It performs gradient boosting and parallezation of tree construction. I have employed this powerful algorithm to predict activity of OU-SALT MS targets.

[1]https://www.kaggle.com/

B.2 *XGBoost*-kPCA model

Before applying the *XGBoost* algorithm with kPCA inputs, it is necessary to optimize predictor variables and tune the algorithm.

B.2.1 Predictor variable optimization

The *XGBoost*-kPCA model predicts preferably on non-logarithmic values. Table B.1 details model performance when each predictor variable is individually removed from the model. Removing $R_G - B_G$, G_V, d_{BJ} and $T_{\mathrm{eff,\,G}}$ causes M1 to weaken to $R^2_{\mathrm{a}} = 0.24 - 0.40$ and BIC $= -359$ to -378. The prediction success of M2 also decreases with the removal of these predictor variables with the exception of $R_G - B_G$, where it marginally increases by Δ BIC $= 1$. Given the detrimental impact of removing this variable from M1, it is retained in both models. The addition of other stellar parameters $M_\star$, $R_\star$, [Fe/H] and $\log g$ does not strengthen the model—the inclusion of these parameters is not justified. The input variables for the *XGBoost*-kPCA models are thus $R_G - B_G$, G_V, d, $T_{\mathrm{eff,\,G}}$ and M_P/a in 4 PCs for M1, and $R_G - B_G$, G_V, d and $T_{\mathrm{eff,\,G}}$ in 3 PCs for M2. In contrast to the OLS-kPCA model, d is included.

	M1					**M2**				
Removed	ρ	R^2_{a}	RMSE	BIC	σ_{res}[%]	ρ	R^2_{a}	RMSE	BIC	σ_{res}[%]
None	0.75	0.52	0.075	-396	67	0.54	0.24	0.095	-363	85
$R_G - B_G$	0.65	0.40	0.084	-378	75	0.55	0.24	0.095	-364	85
G_V	0.52	0.24	0.095	-359	85	0.42	0.11	0.103	-351	92
$\log(d_{BJ})$	0.60	0.35	0.088	-372	79	0.55	0.27	0.093	-366	83
$T_{\mathrm{eff\,G}}$	0.63	0.37	0.086	-375	77	0.38	0.08	0.104	-348	93
Added										
$\log M_\star$	0.62	0.37	0.087	-374	77	0.45	0.14	0.101	-354	90
$\log R_\star$	0.50	0.23	0.095	-359	85	0.26	-0.06	0.112	-337	100
[Fe/H]	0.46	0.18	0.099	-353	88	0.42	0.10	0.103	-350	92
$\log g$	0.57	0.30	0.091	-366	81	0.35	0.03	0.107	-344	96

TABLE B.1: Success metrics when each predictor variable is individually removed from *XGBoost*-kPCA M1 and M2 respectively. The top row details success scores when all variables are included. In the lower half, additional stellar parameters—mass ($M_\star$), radius ($R_\star$), metallicity [Fe/H] and density ($\log g$)—are added to the models to test whether their inclusion is justified.

B.2.2 Tuning *XGBoost*

It is advisable to tune machine learning algorithms to maximise performance. To be clear on terminology: *parameters* are properties of the training data that are learnt during training. They are internal to the model and their values may be estimated from data. In contrast, *hyperparameters* are configurations specified prior to training that aim to help estimate model parameters. They are set before the model begins.

The following *XGBoost* hyperparameters may be specified prior to training:

objective defines the optimization objective to be calculated at each node by defining the loss function to be minimized. My objective is to fit a linear regression;

colsample_bytree is the fraction of columns to be randomly sampled in each tree;

learning_rate makes the model more robust by shrinking the weights at each step;

max_depth is the maximum depth of the tree. This controls over-fitting as more depth allows the model to learn relations specific to the sample;

n_estimators is the number of boosted trees to fit;

sub_sample denotes the fraction of observations to be randomly sampled for each tree—lower values make the algorithm more conservative and may lead to underfitting; and

gamma is the minimum loss reduction required to split a node.

The hyperparameter values specified to optimize the *XGBoost* algorithm to predict OU-SALT MS host activity are listed in Table B.2. Of note, the depth of the tree is '2' for both models, which is lower than default settings. Fewer estimators are required to optimize M2.

Parameter	M1	M2
Objective	reg:linear	reg:linear
Colsample bytree	1.0	1.0
Learning rate	0.13	0.09
Maximum depth	2	2
N estimators	85	60
Sub-sample	1.0	0.8
Gamma	0	0
Verbose	True	True

TABLE B.2: The hyperparameter values used to optimize *XGBOOST* algorithm.

B.2.3 Cross validation

A strategy is required to validate, quantify and compare the success of *XGBoost*-kPCA models. Cross validation is a technique for assessing the accuracy of a predictive models that involves withholding a portion of the dataset to later test model effectiveness (Stone, 1977). The strategy is commonly used in machine learning, where some data is used to *train* the model and other data is used to *test* the model. In k-fold cross-validation, data are divided into k equal subsets. The regression is performed k times, each time with a different subset removed from the training data. Test statistics are collected for each regression to construct a distribution. All observations are used $k-1$ times for training and once for validation. It is common to set $k = 10$.

Leave-one out (LOO) cross validation involves setting $k = N$, so each prediction is based on a training set of $N-1$ observations. This exhaustive form of cross validation is computationally expensive but maximizes the predictive power of data. I employ the LOO strategy to predict activity with the *XGBoost*-kPCA models. This makes model assessment slightly more challenging as, whereas each set of k-folded predictions may be plotted against measured activity to produce test statistics, each LOO prediction generates only one data point. As such, the predictions are collected into a single array, which may be plotted against measured activity and assessed.

B.3 Results

Figure B.2 shows the LOO activity predictions of M1 and M2 plotted against measured activity, with residuals plotted to the right. Predictions made by the OLS-kPCA are also shown for ease of comparison. The residual plot displays less scatter than the OLS-kPCA model at lower values. The residual distributions are presented in Figure B.3. The widths of the distributions are 67 % and 84 % of respective measured activity distributions. This marks improved performance by M1 (from $\sigma_{\text{res}} = 68\,\%$) but weaker performance by M2 (compared to $\sigma_{\text{res}} = 78\,\%$). The mean values of the distributions (2.88×10^{-3} and 6.59×10^{-3}) indicate modest deviations from 0.

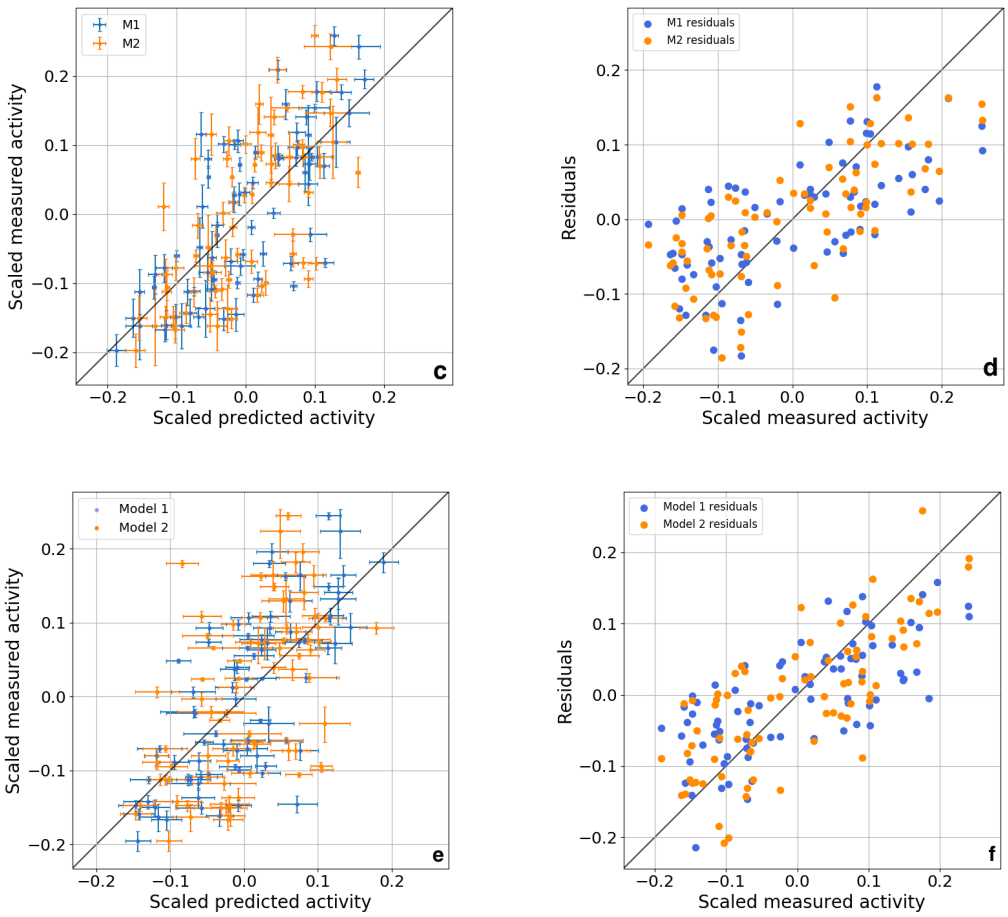

FIGURE B.2: Plots of: (c) OLS-kPCA predictions with 1 σ error bars plotted against scaled measured activity; (d) OLS-kPCA residual between predicted and measured activity; (e) XGBoost-kPCA predictions with 1 σ error bars plotted against scaled measured activity; and (f) XGBoost-kPCA residual between predicted and measured activity.

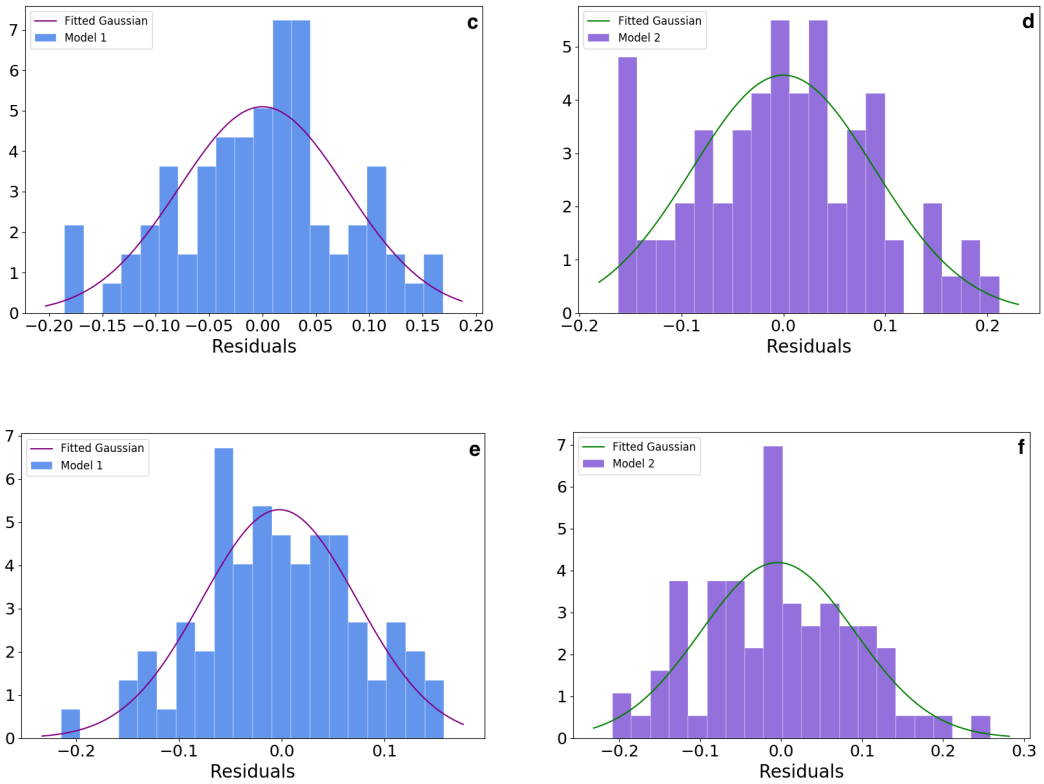

FIGURE B.3: Residuals of predicted and measured activity for: (c) OLS-PCA M1; (d) OLS-PCA M2; (e) OLS-kPCA M1; and (f) OLS-kPCA M2. Residual values are shown on the x axis and frequency on the y axis.

Model success metrics are presented in Table B.3, with distributions plotted in Figure B.4. M1 explains 51 % of variance compared to only 23 % by M2. M1's BIC score is lower by 33, indicating the model is strongly preferred. The residual distribution is 14 % narrower for M1, while the correlation between predicted and measured activity is 17 % stronger. M1 and M2 perform more poorly than the corresponding OLS-kPCA models.

The $|S_{1-2}|$ values range between $1.91 - 2.16\,\sigma$ for bootstrapped distributions and $|S_{1-2}| = 2.73 - 2.87\,\sigma$ for permutation distributions. The latter indicates that there is a $0.4 - 0.6$ % chance that M1 explains the same variance as M2. Permuted null distributions and corresponding p-values between $p = 9.25 \times 10^{-11}$ and 2.01×10^{-5} indicate very low probabilities that these separations are due to chance. This confirms the *XGBoost*-kPCA models effectively predicts activity. The combined distributions are similar to the perturbed distributions, and shift towards the bootstrapped mean in some instances. For example, the mean R^2_a value for M1 increases by 0.01, while the mean BIC value decreases by 4. Separation values decrease from the perturbed models but remain above $|S_{1-2}| = 2\,\sigma$. Combining perturbation and bootstrap methods moves the distributions closer together than for purely permuted separations.

B.4 Summary

The undoubtedly powerful algorithm *XGBoost* has not predicted the activity of OU-SALT MS hosts with the same level of success as the OLS-kPCA model. The *XGBoost*-kPCA M1 and M2 bootstrapped models explain 51 % and 23 % of variance respectively—down 1 % and 18 % from the OLS-kPCA model. While separations between M1 and M2 success metric distributions exceed $2\,\sigma$, this may be because the *XGBoost* algorithm—a big data cruncher—does not cope well with the limited input variables of M2. When tested on smaller sub-samples, results are unstable. Although the *XGBoost*-kPCA model has performed less favourably, it may produce improved prediction success for a larger sample size. I have incorporated the algorithm into a script where it, and any regression algorithm in the *Scikit-learn* library, may be conveniently used to analyse OU-SALT data.

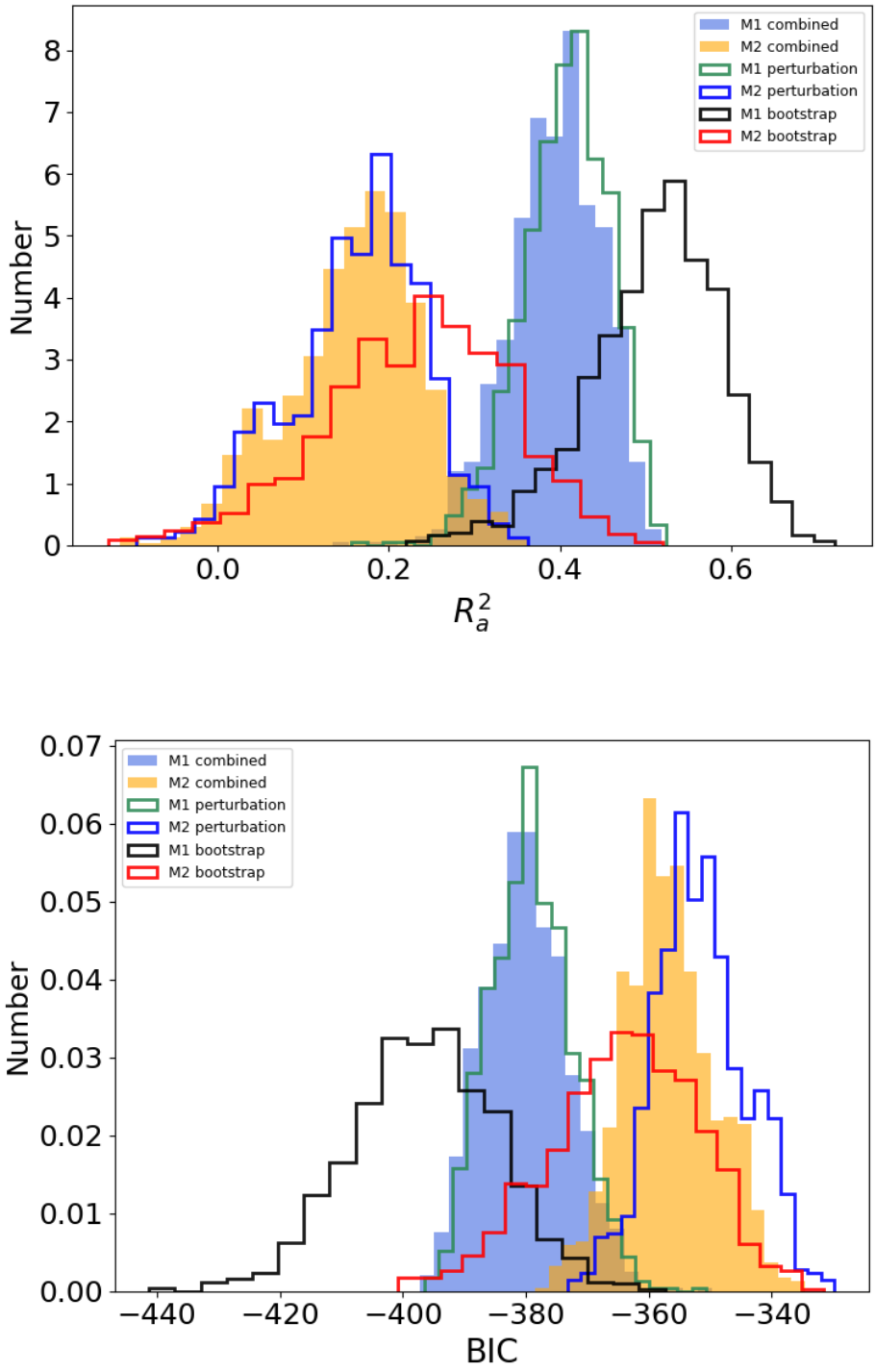

FIGURE B.4: R_a^2 and BIC distributions for *XGBoost*-kPCA Models 1 & 2. Each metric is presented in a separate panel (see *x*-axis labels). Bootstrapped distributions are outlined in black and red, permuted distributions are filled purple, and perturbed distributions are filled blue and orange (see legends). Histogram counts are normalized.

	ρ	R^2_a	RMSE	BIC	$\sigma_{res}[\%]$
Bootstrap					
M1					
μ_{BS1}	0.74	0.51	0.075	-397	67
σ_{BS1}	0.057	0.078	0.006	11.9	5.00
M2					
μ_{BS2}	0.53	0.23	0.095	-364	84
σ_{BS2}	0.084	0.108	0.007	12.0	6.34
S_{1-2}	2.04	2.09	-2.16	-1.91	-2.14
p_{1-2}	0.04	0.04	0.03	0.06	0.03
Perturbation method					
M1					
μ_{PM1}	0.66	0.41	0.083	-380	75
σ_{PM1}	0.039	0.050	0.004	6.8	3.17
M2					
μ_{PM2}	0.48	0.17	0.099	-352	89
σ_{PM2}	0.068	0.074	0.004	7.0	3.91
S_{1-2}	2.24	2.73	-2.82	-2.87	-2.81
p_{1-2}	0.025	0.006	0.005	0.004	0.005
Combined method					
M1					
μ_{PM1}	0.64	0.42	0.082	-384	73
σ_{PM1}	0.087	0.106	0.007	13.1	5.97
M2					
μ_{PM2}	0.41	0.11	0.101	-353	90
σ_{PM2}	0.088	0.099	0.006	9.4	5.32
S_{1-2}	2.16	2.48	-2.67	-2.27	-2.66
p_{1-2}	0.03	0.01	0.01	0.02	0.01
Permutation test					
μ_n	0.010	-0.59	0.137	-301	123
σ_n	0.114	0.171	0.007	8.8	6.67
M1					
S_{1-n}	5.85	6.04	-6.63	-6.48	-6.68
p_{1-n}	4.93e-09	1.53e-09	3.38e-11	9.25e-11	2.34e-11
M2					
S_{2-n}	3.81	4.25	-4.12	-4.26	-4.18
p_{2-n}	1.40e-04	2.12e-05	3.79e-05	2.01e-05	2.93e-05

TABLE B.3: Success metrics as described in the caption for Table 6.7 for *XGBoost*-kPCA model predictions.

Bibliography

Adams F. C., 2011, ApJ, 730, 27

Adams E. R., Jackson B., Endl M., 2016, AJ, 152, 47

Ahuir J., Brun A. S., Strugarek A., 2020, A&A, 365

Aizerman M. A., Braverman E. M., Rozoner L. I., 1964, Automation and Remote Control, 25, 821

Akaike H., 1974, IEEE Transactions on Automatic Control, 19, 716

Albrecht S., Winn J. N., Butler R. P., Crane J. D., Shectman S. A., Thompson I. B., Hirano T., Wittenmyer R. A., 2012, ApJ, 744

Allan A., Vidotto A. A., 2019, MNRAS, 490, 3760

Allart R., et al., 2018, Science, 362, 1384

Allart R., et al., 2019, A&A, 623

Allen C. W., 1973, Astrophysical Quantities, 3 edn. University of London, Athlone Press, London

Anderson, D. R. et al., 2017, A&A, 604, A110

Anderson D. R., et al., 2011, MNRAS, 416, 2108

Andretta V., Busà I., Gomez M. T., Terranegra L., 2005, A&A, 430, 669

Anglada-Escudé G., et al., 2016, ApJ, 830, 74

Angus R., et al., 2019, AJ, 158, 173

Apostol T. M., 1976, Introduction to Analytic Number Theory. Springer, Berlin, Heidelberg

Arcangeli J., et al., 2019, A&A, 625, A136

Arenou F., et al., 2018, A&A, 616

Armstrong D. J., Meru F., Bayliss D., Kennedy G. M., Veras D., 2019, ApJ, 880, L1

Ayres T. R., 2010, ApJS, 187, 149

Bailer-Jones C. A. L., Rybizki J., Fouesneau M., Mantelet G., Andrae R., 2018, AJ, 156, 11

Bakos G., Noyes R. W., Kovacs G., Stanek K. Z., Sasselov D. D., Domsa I., 2004, PASP, 116, 266

Bakos G. Á., et al., 2013, PASP, 125, 154

Baliunas S. L., et al., 1995, ApJ, 438, 269

Baluev R. V., 2008, MNRAS, 385, 1279

Baraffe I., Selsis F., Chabrier G., Barman T. S., Allard F., Hauschildt P. H., Lammer H., 2004, A&A, 419, L13

Baraffe I., Chabrier G., Barman T. S., Selsis F., Allard F., Hauschildt P. H., 2005a, A&A, 436, L47

Baraffe I., Alibert Y., Chabrier G., Benz W., 2005b, A&A, 450, 1221

Barnes S. A., 2003, ApJ, 586, 464

Barnes S. A., 2007, ApJ, 669, 1167

Barnes J. R., Haswell C. A., Staab D., Anglada-Escudé G., 2016, MNRAS, 462, 1012

Barnes J. R., et al., 2019, Nature Astronomy, 4, 419

Barros, S. C. C. et al., 2011, A&A, 525, A54

Barros, S. C. C. et al., 2016, A&A, 593, A113

Basturk O., Yalcinkaya S., Keten B., 2019, preprint (arXiv:1911.08330)

Batygin K., Bodenheimer P. H., Laughlin G. P., 2016, ApJ, 829, 114

Bayliss D., et al., 2018, MNRAS, 475, 4467

Beaugé C., Nesvorný D., 2013, ApJ, 763, 12

Béky B., et al., 2011, ApJ, 734, 109

Belenkaya E. S., Khodachenko M. L., Alexeev I. I., 2014, in Lammer H., Khodachenko M., eds, Astrophysics and Space Science Library, Vol. 411, Characterizing Stellar and Exoplanetary Environments. Springer, Cham, pp 239–249, doi:10.1007/978-3-319-09749-7_12

Bell T. J., Cowan N. B., 2018, ApJ, 857, L20

Ben-Jaffel L., Ballester G. E., 2013, A&A, 553, A52

Benz W., Ehrenreich D., Isaak K., 2018, CHEOPS: CHaracterizing ExOPlanets Satellite. Springer, Cham, doi:10.1007/978-3-319-30648-3_84-1

Bishop C. M., 2015, Pattern Recogniton and Machine Learning, 1 edn. Springer-Verlag, New York

Bisikalo D., Kaygorodov P., Ionov D., Shematovich V., Lammer H., Fossati L., 2013, ApJ, 764, 19

Bochinski J. J., Haswell C. A., Marsh T. R., Dhillon V. S., Littlefair S. P., 2015, ApJ, 800, L21

Boley A. C., Contreras A. P. G., Gladman B., 2016, ApJ, 817, L17

Boro Saikia S., et al., 2018, A&A, 616, A108

Borucki W. J., et al., 2010, Science, 327, 977

Boser B. E., Guyon I. M., Vapnik V. N., 1992, in Proceedings of the Fifth Annual Workshop on Computational Learning Theory. ACM, New York, pp 144–152, doi:10.1145/130385.130401

Bouchy F., et al., 2005, A&A, 444, L15

Bourrier V., et al., 2018, A&A, 620, A147

Bouvier J., Forestini M., Allain S., 1997, A&A, 326, 1023

Bouvier J., et al., 2006, A&A, 463, 1017

Brandenburg A., Giampapa M. S., 2018, ApJ, 855, L22

Brandenburg A., Saar S. H., Turpin C. R., 1998, ApJ, 498, L51

Brown T. M., 2001, ApJ, 553, 1006

Brown T. M., et al., 2013, PASP, 125, 1031

Brown W. R., Lattanzi M. G., Kenyon S. J., Geller M. J., 2018, ApJ, 866, 39

Brun A. S., Browning M. K., Dikpati M., Hotta H., Strugarek A., 2015a, Space Science Reviews, 196, 101

Brun A. S., García R. A., Houdek G., Nandy D., Pinsonneault M., 2015b, Space Science Reviews, 196, 303

Buccino A. P., Mauas P. J. D., 2008, A&A, 483, 903

Buchholz B., Ulmschneider P., Cuntz M., 1998, ApJ, 494, 700

Buckley D. A. H., Swart G. P., Meiring J. G., 2006, in Stepp L. M., ed., Vol. 6267, Ground-based and Airborne Telescopes. SPIE, pp 333–347, doi:10.1117/12.673750

Burke E. W., Rolland W. W., Boy W. R., 1970, Journal of the Royal Astronomical Society of Canada, 64, 353

Burrows A., Lunine J., 1995, Nature, 378, 333

Busuttil R., 2017, PhD thesis, The Open University

Buzasi D., Lezcano A., Preston H. L., 2016, Journal of Space Weather and Space Climate, 6, A38

Cabrera J., et al., 2015, A&A, 579

Camenzind M., 1990, in Klare G., ed., Reviews in Modern Astronomy, Vol. 3, Accretion and Winds. pp 234–265, doi:10.1007/978-3-642-76238-3_17

Canto Martins B. L., das Chagas M. L., Alves S., Leão I. C., de Souza Neto L. P., de Medeiros J. R., 2011, A&A, 530, A73

Carrera D., Ford E. B., Izidoro A., 2019, MNRAS, 486, 3874

Carroll-Nellenback J., Frank A., Liu B., Quillen A. C., Blackman E. G., Dobbs-Dixon I., 2017, MNRAS, 466, 2458

Casasayas-Barris N., et al., 2018, A&A, 616, A151

Catala C., Donati J. F., Shkolnik E., Bohlender D., Alecian E., 2006, MNRAS, 374, L42

Cauley P. W., Redfield S., Jensen A. G., 2017, AJ, 153, 217

Cauley P. W., Shkolnik E. L., Llama J., Bourrier V., Moutou C., 2018, AJ, 156, 262

Cauley P. W., Shkolnik E. L., Llama J., Lanza A. F., 2019a, Nature Astronomy, 3, 1128

Cauley P. W., Shkolnik E. L., Ilyin I., Strassmeier K. G., Redfield S., Jensen A., 2019b, AJ, 157, 69

Chamberlain J. W., 1963, Planetary and Space Science, 11, 901

Charbonneau P., 2010, Living Reviews in Solar Physics, 7, 3

Charbonneau D., Brown T. M., Latham D. W., Mayor M., 2000, ApJ, 529, L45

Charbonneau D., et al., 2005, ApJ, 626, 523

Chen T., Guestrin C., 2016, preprint, pp 785–794 (arXiv:1603.02754)

Chen H., Rogers L. A., 2016, ApJ, 831, 180

Chiang E., Laughlin G., 2013, MNRAS, 431, 3444

Chmielewski Y., 2000, A&A, 353, 666

Christiansen J. L., et al., 2017, AJ, 154, 122

Clarke D., 2002, A&A, 386, 763

Cohen O., Drake J. J., Kashyap V. L., Saar S. H., Sokolov I. V., Manchester W. B., Hansen K. C., Gombosi T. I., 2009, ApJ, 704

Cohen O., Kashyap V. L., Drake J. J., Sokolov I. V., Garraffo C., Gombosi T. I., 2011, ApJ, 733, 67

Collins K. A., Kielkopf J. F., Stassun K. G., Hessman F. V., 2017, AJ, 153, 77

Coughlin J. L., et al., 2015, ApJS, 224, 12

Crawford S. M., et al., 2010, in Silva D. R., Peck A. B., Soifer B. T., eds, Vol. 7737, Observatory Operations: Strategies, Processes, and Systems III. SPIE, pp 550 – 561, doi:10.1117/12.857000

Cubillos P. E., Fossati L., Koskinen T., Young M. E., Salz M., France K., Sreejith A. G., Haswell C. A., 2020, AJ, 159, 2020

Cumming A., Butler R. P., Marcy G. W., Vogt S. S., Wright J. T., Fischer D. A., 2008, PASP, 120, 531

Cuntz M., Saar S. H., Musielak Z. E., 2000, ApJ, 533, L151

Curran P. A., 2014, preprint (arXiv:1411.3816)

Czesla S., Salz M., Schneider P. C., Schmitt J. H. M. M., 2013, A&A, 560, A17

D'Angelo C. V., Vidotto A. A., Esquivel A., Sgró M. A., Koskinen T., Fossati L., 2019, preprint (arXiv:1912.03778)

Dai F., Masuda K., Winn J. N., 2018, ApJ, 864, L38

Daley-Yates S., Stevens I. R., 2018, MNRAS, 479, 1194

Davies R. B., 1987, Biometrika, 74, 33

De Pater I., Lissauer J. J., 2015, Planetary Atmospheres, 2 edn. Cambridge University Press, Cambridge, doi:10.1017/CBO9781316165270.005

Debrecht A., Carroll-Nellenback J., Frank A., Fossati L., Blackman E. G., Dobbs-Dixon I., 2018, MNRAS, 478, 2592

Deming D., Seager S., Richardson L. J., Harrington J., 2005, Nature, 434, 740

Demory B. O., et al., 2011, ApJ, 735, L12

Demory B. O., et al., 2013, ApJ, 776, L25

Désert J. M., et al., 2011, ApJS, 197, 14

Doherty J. P. J., Haswell C. A., Barnes J. R., Staab D., Fossati L., 2019, in European Planetary Sciences Congress—Division for Planetary Sciences Joint Meeting 2019. pp 1831–4, https://meetingorganizer.copernicus.org/EPSC-DPS2019/EPSC-DPS2019-1831-4.pdf

Donati J. F., Landstreet J., 2009, ARA&A, 47, 333

Dotter A., Chaboyer B., Jevremović D., Kostov V., Baron E., Ferguson J. W., 2008, ApJS, 178, 89

Dressing C. D., et al., 2015, ApJ, 800, 135

Duncan D. K., et al., 1991, ApJS, 76, 383

Dwivedi N. K., et al., 2019, MNRAS, 487, 4208

Dworetsky M. M., 1983, MNRAS, 203, 917

Eberhard G., Schwarzschild K., 1913, ApJ, 38, 292

Efron B., 1979, The Annals of Statistics, 501, 1979

Ehrenreich D., Désert J.-M., 2011, A&A, 529, A136

Ehrenreich D., et al., 2015, Nature, 522, 459

Eker Z., et al., 2015, AJ, 149, 131

Erkaev N. V., Kulikov Y. N., Lammer H., Selsis F., Langmayr D., Jaritz G. F., Biernat H. K., 2007, A&A, 472, 329

Esposito M., et al., 2017, A&A, 601, A53

Fares R., et al., 2009, MNRAS, 398, 1383

Fares R., et al., 2010, MNRAS, 406, 409

Fares R., et al., 2012, MNRAS, 423, 1006

Feigelson E. D., Babu G. J., 2012, Modern Statistical Methods for Astronomy: With R Applications. Cambridge University Press, Cambridge, doi:10.1017/CBO9781139015653

Figueira P., Oshagh M., Adibekyan V. Z., Santos N. C., 2014, A&A, 572, A51

Fischer C., Saur J., 2019, ApJ, 872, 113

Fleming B. T., et al., 2018, Journal of Astronomical Telescopes, Instruments, and Systems, 4, 1

Folsom C. P., Fionnagáin D. Ó., Fossati L., Vidotto A. A., Moutou C., Petit P., Dragomir D., Donati J.-F., 2020, A&A, 633, A48

Fontanive C., Rice K., Bonavita M., Lopez E., Mužić K., Biller B., 2019, MNRAS, 485, 4968

Fortney J. J., et al., 2011, ApJS, 197, 9

Fossati L., et al., 2010, ApJ, 714, L222

Fossati L., Ryabchikova T., Shulyak D. V., Haswell C. A., Elmasli A., Pandey C. P., Barnes T. G., Zwintz K., 2011, MNRAS, 417, 495

Fossati L., Ayres T. R., Haswell C. A., Bohlender D., Kochukhov O., Flöer L., 2013, ApJ, 766, L20

Fossati L., Ingrassia S., Lanza a. F., 2015a, ApJ, 812, L35

Fossati L., France K., Koskinen T., Juvan I. G., Haswell C. A., Lendl M., 2015b, ApJ, 815, 118

Fossati L., et al., 2017a, A&A, 598, A90

Fossati L., et al., 2017b, A&A, 601, A104

Fossati L., Koskinen T., France K., Cubillos P. E., Haswell C. A., Lanza A. F., Pillitteri I., 2018a, AJ, 155, 113

Fossati L., Koskinen T., Lothringer J. D., France K., Young M. E., Sreejith A. G., 2018b, ApJ, 868, L30

Foukal P., 2018, preprint (arXiv:1810.06558)

France K., et al., 2016, ApJ, 820, 89

France K., Arulanantham N., Fossati L., Lanza A. F., Loyd R. O. P., Redfield S., Schneider P. C., 2018, ApJS, 239, 16

Freeman J. B., Dale R., 2013, Behaviour Research Methods, 45, 83

Fressin F., et al., 2013, ApJ, 766

Freund Y., Schapire R. E., 1997, Journal of Computer System Sciences, 55, 119–139

Friedman J. H., 2001, The Annals of Statistics, 29, 1189

Fulton B. J., Petigura E. A., 2018, AJ, 156, 264

Fulton B. J., et al., 2017, AJ, 154, 109

Gallet F., Delorme P., 2019, A&A, 626

Gandolfi D., et al., 2010, A&A, 524

Gandolfi D., et al., 2015, A&A, 576

Gao P., Marley M. S., Zahnle K., Robinson T. D., Lewis N. K., 2017, AJ, 153, 139

García Muñoz A., Schneider P. C., 2019, ApJ, 884, L43

García R. A., et al., 2014, A&A, 572

Gardner J. P., et al., 2006, Space Science Reviews, 123, 485

Gaudi B. S., 2005, ApJ, 628, L73

Gaudi B. S., et al., 2017, Nature, 546, 514

Getman K. V., Broos P. S., Salter D. M., Garmire G. P., Hogerheijde M. R., 2011, ApJ, 730, 6

Getman K. V., Broos P. S., Kóspál Á., Salter D. M., Garmire G. P., 2016, AJ, 152, 188

Giampapa M. S., 2015, Proceedings of the International Astronomical Union, 11, 365

Gibson N. P., et al., 2020, MNRAS, 493, 2215

Gillon M., et al., 2011, A&A, 533

Gillon M., et al., 2012, A&A, 542, A4

Gillon M., et al., 2014, A&A, 562, L3

Gillon M., et al., 2017, Nature, 542, 456

Ginzburg S., Schlichting H. E., Sari R., 2016, ApJ, 825, 29

Ginzburg S., Schlichting H. E., Sari R., 2018, MNRAS, 476, 759

Girardi L., Grebel E. K., Odenkirchen M., Chiosi C., 2004, A&A, 422, 205

Girardi L., et al., 2008, PASP, 120, 583

Gomes Da Silva J., Santos N. C., Boisse I., Dumusque X., Lovis C., 2014, A&A, 566

Gratier P., et al., 2017, A&A, 599, A100

Gray R. O., Corbally C. J., Garrison R. F., McFadden M. T., Robinson P. E., 2003, AJ, 126, 2048

Gray R. O., Corbally C. J., Garrison R. F., McFadden M. T., Bubar E. J., McGahee C. E., O'Donoghue a. a., Knox E. R., 2006, AJ, 132, 161

Green S., Jones M., 2015, An Introduction to the Sun and Stars, 2 edn. Cambridge University Press, Cambridge

Guenther E. W., et al., 2017, A&A, 608

Guillot T., Burrows A., Hubbard W. B., Lunine J. I., Saumon D., 1996, ApJ, 459, L35

Guinan E. F., Engle S. G., 2008, Proceedings of the International Astronomical Union, 4, 395–408

Guo J. H., 2011, ApJ, 733, 98

Gupta A., Schlichting H. E., 2019, MNRAS, 487, 24

Gupta A., Schlichting H. E., 2020, MNRAS, 493, 792

Hall J. C., 2005, Living Reviews in Solar Physics, 5

Hamer J. H., Schlaufman K. C., 2019, AJ, 158, 190

Hansen M., Peng Oh S., 2006, MNRAS, 367, 979

Hartigan J. A., Hartigan P. M., 1985, The Annals of Statistics, 13, 70

Hartman J. D., 2010, ApJ, 717, L138

Hartman J. D., et al., 2015, AJ, 149, 166

Haswell C. A., 2010, Transiting Exoplanets, 1 edn. Cambridge University Press, Cambridge

Haswell C. A., et al., 2012, ApJ, 760, 79

Haswell C. A., et al., 2019, Nature Astronomy, 4, 408

Haukoos J. S., Lewis R. J., 2005, Academic Emergency Medicine, 12, 360

Hebb L., et al., 2009, ApJ, 693, 1920

Hébrard G., des Étangs A. L., Vidal-Madjar A., Désert J. M., Ferlet R., 2003, preprint (arXiv:0312384)

Hébrard G., et al., 2013, A&A, 549

Heller R., Rodenbeck K., Bruno G., 2019, A&A, 624

Hellier C., et al., 2011, A&A, 535, L7

Hellier C., et al., 2012, MNRAS, 426, 739

Hellier C., et al., 2014, MNRAS, 440, 1982

Henry T. J., Soderblom D. R., Donahue R. a., Baliunas S. L., 1996, AJ, 111, 439

Henry G. W., Baliunas S. L., Donahue R. a., Fekel F. C., Soon W. H., 2000, ApJ, 531, 415

Hernán-Obispo M., Gálvez-Ortiz M. C., Anglada-Escudé G., Kane S. R., Barnes J. R., de Castro E., Cornide M., 2010, A&A, 512, A45

Hernán-Obispo M., et al., 2015, A&A, 576, A66

Hertzsprung E., 1923, Bulletin of the Astronomical Institutes of the Netherlands, 2, 15

Hillenbrand L., Isaacson H., Marcy G., Barenfeld S., Fischer D., Howard A., 2014, preprint, pp 759–766 (arXiv:1408.3475)

Hoeijmakers H. J., et al., 2019, A&A, 627

Horne K., 1986, PASP, 98, 609

Howard A. W., et al., 2013, Nature, 503, 381

Howe A. R., Adams F. C., Meyer M. R., 2020, ApJ, 894, 130

Hoyle F., Wilson O. C., 1958, ApJ, 128, 604

Hubbard W. B., Hattori M. F., Burrows A., Hubeny I., Sudarsky D., 2007, Icarus, 187, 358

Ibanoglu C., Soydugan F., Soydugan E., Dervisoglu A., 2006, MNRAS, 373, 435

Ida S., Lin D. N. C., 2004, ApJ, 604, 388

Indriolo N., Hobbs L. M., Hinkle K. H., McCall B. J., 2009, ApJ, 703, 2131

Ip W.-H., Kopp A., Hu J.-H., 2004, ApJ, 602, L53

Iro N., Deming L. D., 2010, ApJ, 712, 218

Irzik G., Meyer E., 1987, The University of Chicago Press Journals, 54, 495

Isaacson H., Fischer D., 2010, ApJ, 725, 875

Jaritz G. F., Endler S., Langmayr D., Lammer H., Grießmeier J. M., Erkaev N. V., Biernat H. K., 2005, A&A, 439, 771

Jeans J. H., 1925, MNRAS, 85, 912

Jenkins J. S., et al., 2011, A&A, 531, A8

Jin S., Mordasini C., Parmentier V., van Boekel R., Henning T., Ji J., 2014, ApJ, 795, 65

Joergens V., Müller A., Reffert S., 2010, A&A, 521

Johnson J. A., et al., 2012, AJ, 143, 111

Johnstone C. P., Khodachenko M. L., Lüftinger T., Kislyakova K. G., Lammer H., Güdel M., 2019, A&A, 624

Jolliffe I., 2002, Encyclopedia of Statistics in Behavioral Science, 3, 1580

Jontof-Hutter D., Lissauer J. J., Rowe J. F., Fabrycky D. C., 2014, ApJ, 785, 15

Jowett G. H., Fisher R. A., 1956, Applied Statistics, 5, 68

Judge P. G., Saar S. H., 2007, ApJ, 663, 643

Kashyap V., Drake J., Saar S., 2008, ApJ, 687

Kawaler S. D., 1988, ApJ, 333, 236

Kendall M. G., 1938, Biometrika, 30, 81

King J. R., Villarreal A. R., Soderblom D. R., Gulliver A. F., Adelman S. J., 2003, AJ, 125, 1980

Kipping D. M., Spiegel D. S., 2011, MNRAS, 417, L88

Knutson H. a., Howard A. W., Isaacson H., 2010, ApJ, 720, 1569

Kobulnicky H. A., Nordsieck K. H., Burgh E. B., Smith M. P., Percival J. W., Williams T. B., O'Donoghue D., 2003, in Iye M., Moorwood A. F. M., eds, Vol. 4841, Instrument Design and Performance for Optical/Infrared Ground-based Telescopes. SPIE, pp 1634 – 1644, doi:10.1117/12.460315

Koenigl A., 1991, ApJ, 370, L39

Komacek T. D., Showman A. P., 2016, ApJ, 821, 16

Kraft R. P., 1967, ApJ, 150, 551

Krause F., Raedler K. H., 1980, Mean-field magnetohydrodynamics and dynamo theory

Kreidberg L., et al., 2018, AJ, 156, 17

Krejčová T., Budaj J., 2012, A&A, 540, A82

Kubyshkina D., et al., 2018a, A&A, 619

Kubyshkina D., et al., 2018b, ApJ, 866, L18

Kubyshkina D., et al., 2019, ApJ, 879, 26

Kulow J. R., France K., Linsky J., Parke Loyd R. O., 2014, ApJ, 786, 132

Kurokawa H., Nakamoto T., 2014, ApJ, 783, 54

Kutner M., 2005, Applied Linear Statistical Models, 5 edn. McGraw-Hill Irwin, Boston

Lafler J., Kinman T. D., 1965, ApJS, 11, 216

Lam K. W., et al., 2017, A&A, 599

Lammer H., Selsis F., Ribas I., Guinan E. F., Bauer S. J., Weiss W. W., 2003, ApJ, 598, L121

Lanza A. F., 2008, A&A, 487, 1163

Lanza A. F., 2009, A&A, 505, 339

Lanza A. F., 2012, A&A, 544, A23

Lanza A. F., 2013, A&A, 557, A31

Lanza A. F., 2014, A&A, 572, L6

Lavie B., et al., 2017, A&A, 605, L7

Lebreton Y., Montalbán J., 2008, Proceedings of the International Astronomical Union, 4, 419–430

Lecavelier des Etangs A., 2006, A&A, 461, 1185

Lecavelier des Etangs A., Vidal-Madjar A., McConnell J. C., Hébrard G., 2004, A&A, 418, L1

Lecavelier des Etangs A., et al., 2010, A&A, 514, A72

Lee E. J., Chiang E., 2017, ApJ, 842, 40

Léger A., et al., 2004, Icarus, 169, 499

Léger A., et al., 2009, A&A, 506, 287

Leighton R. B., 1959, ApJ, 130, 366

Lendl M., et al., 2014, A&A, 568

Lendl M., Cubillos P., Hagelberg J., Müller A., Juvan I., Fossati L., 2017, A&A, 606, A18

Lépine J., Gregorio-Hetem J., 2003, Open Issues in Local Star Formation, 1 edn. Vol. 299 of the Ap&SS Library, Springer Netherlands, doi:10.1007/1-4020-2600-5

Lindegren L., et al., 2018, A&A, 616

Linsky J. L., 2019, Host Stars and their Effects on Exoplanet Atmospheres, 1 edn. Vol. 955 of Lecture Notes in Physics, Springer, doi:10.1007/978-3-030-11452-7

Linsky J. L., Hunten D. M., Sowell R., Glackin D. L., Kelch W. L., 1979, ApJS, 41, 481

Locci D., Cecchi-Pestellini C., Micela G., 2019, A&A, 624

Lomb N. R., 1976, Ap&SS, 39, 447

Lopez E. D., Fortney J. J., 2013, ApJ, 776, 2

Lopez E. D., Fortney J. J., 2014, ApJ, 792, 1

Lopez E. D., Rice K., 2018, MNRAS, 479, 5303

Lopez-Morales M., 2010, preprint (arXiv:1012.2157)

López-Morales M., Seager S., 2007, ApJ, 667, L191

Lopez E. D., Fortney J. J., Miller N., 2012, ApJ, 761, 59

Lorenzo-Oliveira D., Porto De Mello G. F., Schiavon R. P., 2016, A&A, 594

Lorenzo-Oliveira D., et al., 2018, A&A, 619, A73

Lovis C., et al., 2011a, preprint (arXiv:1107.5325)

Lovis C., et al., 2011b, preprint (arXiv:1107.5325)

Loyd R. O., et al., 2016, ApJ, 824, 102

Loyd R. O. P., Shkolnik E. L., Schneider A. C., Richey-Yowell T., Barman T. S., Peacock S., Pagano I., 2020, ApJ, 890, 23

Lubin D., Tytler D., Kirkman D., 2012, ApJ, 747, L32

Lumb D. H., 2012, Optical Engineering, 51, 011009

Luri X., et al., 2018, A&A, 616, A9

Maciejewski G., et al., 2018, Acta Astronomica, 68, 371

Maggio a., et al., 2015, ApJ, 811, L2

Mamajek E. E., Hillenbrand L. a., 2008, ApJ, 687, 1264

Mancini L., et al., 2014, A&A, 562

Mansfield M., et al., 2018, ApJ, 868, L34

Martin E. L., Basri G., Pavlenko Y., Lyubchik Y., 2002, ApJ, 579, 437

Martins B. L. C., das Chagas M. L., Alves S., Leão I. C., Neto L. P. d. S., de Medeiros J. R., 2011, A&A, 530, A73

Mascareño A. S., Rebolo R., González Hernández J. I., Esposito M., 2015, MNRAS, 452, 2745

Mathis S., Le Poncin-Lafitte C., 2009, A&A, 497, 889

Matsakos T., Königl A., 2016, ApJ, 820, L8

Matsakos T., Uribe A., Königl A., 2015, A&A, 578, A6

Mayor M., Queloz D., 1995, Nature, 378, 355

Mazeh T., Holczer T., Faigler S., 2016, A&A, 589

McCormac J., et al., 2020, MNRAS, 493, 126

McCully C., Volgenau N. H., Harbeck D.-R., Lister T. A., Saunders E. S., Turner M. L., Siiverd R. J., Bowman M., 2018, in Guzman J. C., Ibsen J., eds, Vol. 10707, Software and Cyberinfrastructure for Astronomy V. SPIE, pp 141 – 149, doi:10.1117/12.2314340

McQuillan A., Aigrain S., Mazeh T., 2013a, MNRAS, 432, 1203

McQuillan A., Mazeh T., Aigrain S., 2013b, ApJ, 775, L1

McQuillan A., Mazeh T., Aigrain S., 2014, ApJS, 211

Mengel M. W., Fares R., Marsden S. C., Carter B. D., Jeffers S. V., Petit P., Donati J. F., Folsom C. P., 2016, MNRAS, 459, 4325

Meyer M., Amara A., Reggiani M., Quanz S. P., 2018, A&A, 612, L3

Middelkoop F., 1982, A&A, 107, 31

Miesch M. S., 2012, Philosophical Transactions of the Royal Society of London Series A, 370, 3049

Miller B. P., Gallo E., Wright J. T., Dupree A. K., 2012, ApJ, 754, 137

Miller B. P., Gallo E., Wright J. T., Pearson E. G., 2015, ApJ, 799, 163

Mittag M., Schmitt J. H. M. M., Schröder K.-P., 2013, A&A, 549, A117

Moffatt H. K., 1978, Magnetic Field Generation in Electrically Conducting Fluids. Cambridge Monographs on Mechanics and Applied Mathematics, Cambridge University Press

Montes D., Fernandez-Figueroa M. J., De Castro E., Cornide M., Latorre A., Sanz-Forcada J., 2000, A&AS, 146, 103

Montes D., López-Santiago J., Gálvez M. C., Fernández-Figueroa M. J., De Castro E., Cornide M., 2001, MNRAS, 328, 45

Mordasini C., 2020, preprint (arXiv:2002.02455)

Mordasini C., Alibert Y., Benz W., Naef D., 2009, A&A, 501, 1161

Mordasini C., van Boekel R., Mollière P., Henning T., Benneke B., 2016, ApJ, 832, 41

Morris B. M., Mandell A. M., Deming D., 2013, ApJL, 764, L22

Moutou C., et al., 2014, MNRAS, 444, 2783

Murgas F., Pallé E., Zapatero Osorio M. R., Nortmann L., Hoyer S., Cabrera-Lavers a., 2014, A&A, 563, A41

Murray-Clay R. A., Chiang E. I., Murray N., 2009, ApJ, 693, 23

Narain U., Ulmschneider P., 1996, Space Science Reviews, 75, 453

Neveu-VanMalle M., et al., 2014, A&A, 572, A49

Nortmann L., et al., 2018, Science, 362, 1388

Nowak G., et al., 2017, AJ, 153, 131

Noyes R. W., Hartmann L. W., Baliunas S. L., Duncan D. K., Vaughan A. H., 1984, ApJ, 279, 763

O'Keefe J. D., Ahrens T. J., 1985, Icarus, 62, 604

Ofir A., Dreizler S., Zechmeister M., Husser T.-O., 2014, A&A, 561, A103

Oklopčić A., 2019, ApJ, 881, 133

Oklopčić A., Hirata C. M., 2018, ApJ, 855, L11

Öpik E. J., 1963, Geophysical Journal International, 7, 490

Ossendrijver M., 2003, A&ARv, 11, 287

Owen J. E., 2019, Annual Review of Earth and Planetary Sciences, 47, 67

Owen J. E., Jackson A. P., 2012, MNRAS, 425, 2931

Owen J. E., Wu Y., 2013, ApJ, 775, 105

Owen J. E., Wu Y., 2017, ApJ, 847, 29

Pace G., 2013, A&A, 551, L8

Pace G., Pasquini L., 2004, A&A, 426, 1021

Paczynski B., 1971, ARA&A, 9, 183

Parker J. G., 1956, The Journal of the Acoustical Society of America, 28, 672

Parker E. N., 1979, ApJ, 230, 905

Parmentier V., Fortney J. J., Showman A. P., Morley C. V., Marley M. S., 2016, ApJ, 828, 22

Parmentier V., et al., 2018, A&A, 617, A110

Pearson K., Karl 1895, Philosophical Transactions of the Royal Society, 186, 343

Pedregosa F., et al., 2011, Journal of Machine Learning Research, 12, 2825

Penev K., Jackson B., Spada F., Thom N., 2012, ApJ, 751

Penev K., et al., 2016, AJ, 152, 127

Pepper J., et al., 2017, AJ, 153, 215

Peres-Neto P. R., Jackson D. A., Somers K. M., 2005, Computational Statistics and Data Analysis, 49, 974

Perez-Becker D., Chiang E., 2013, MNRAS, 433, 2294

Perryman M., 2018, The Exoplanet Handbook, 2 edn. Cambridge University Press, Cambridge

Petigura E. A., et al., 2017a, AJ, 153, 142

Petigura E. A., et al., 2017b, AJ, 155, 89

Petrovich C., Deibert E., Wu Y., 2019, AJ, 157, 180

Pillitteri I., Günther H. M., Wolk S. J., Kashyap V. L., Cohen O., 2011, ApJ, 741, L18

Pillitteri I., Wolk S. J., Sciortino S., Antoci V., 2014, A&A, 567, A128

Pinhas A., Madhusudhan N., Clarke C., 2016, MNRAS, 463, 4516

Pollacco D., et al., 2006, PASP, 118, 1407

Poppenhaeger K., 2015, EPJ Web of Conferences, 101

Poppenhaeger K., 2016, Proceedings of the International Astronomical Union, 12, 308

Poppenhaeger K., Schmitt J. H. M. M., 2011, ApJ, 735, 59

Poppenhaeger K., Wolk S. J., 2014, A&A, 565, L1

Poppenhaeger K., Robrade J., Schmitt J. H. M. M., 2010, A&A, 515, A98

Poppenhaeger K., Czesla S., Schröter S., Lalitha S., Kashyap V., Schmitt J. H., 2012, A&A, 541

Poppenhaeger K., Schmitt J. H. M. M., Wolk S. J., 2013, ApJ, 773, 62

Popper D. M., 1967, ARA&A, 5, 85

Popper D. M., 1980, ARA&A, 112, 454

Post M., Wolf S., Stock G., 2019, Journal of Chemical Physics, 150, 204110

Press W. H., Rybicki G. B., 1989, ApJ, 338, 277

Price-Whelan A. M., et al., 2018, AJ, 156, 123

Privitera G., Meynet G., Eggenberger P., Vidotto A. A., Villaver E., Bianda M., 2016a, A&A, 591

Privitera G., et al., 2016c, A&A, 593

Privitera G., Meynet G., Eggenberger P., Vidotto A. A., Villaver E., Bianda M., 2016b, A&A, 593

Prusti T., 2011, EAS Publications Series, 45, 9

Pu B., Lai D., 2018, MNRAS, 478, 197

Queloz D., et al., 2001, A&A, 379, 279

Queloz D., et al., 2009, A&A, 506, 303

Quirrenbach A., et al., 2014, in Ramsay S. K., McLean I. S., Takami H., eds, Vol. 9147, Ground-based and Airborne Instrumentation for Astronomy V. SPIE, pp 531 – 542, doi:10.1117/12.2056453

Raftery A. E., 1995, Sociological Methodology, 25, 111

Rappaport S., et al., 2012, ApJ, 752, 1

Rappaport S., Sanchis-Ojeda R., Rogers L. a., Levine A., Winn J. N., 2013, ApJ, 773, L15

Rappaport S., Barclay T., DeVore J., Rowe J., Sanchis-Ojeda R., Still M., 2014, ApJ, 784, 40

Rauer H., et al., 2014, Experimental Astronomy, 38, 249

Rebull L. M., et al., 2016a, AJ, 152, 113

Rebull L. M., et al., 2016b, AJ, 152, 114

Redfield S., Linsky J. L., 2008, ApJ, 673, 283

Reiners A., Mohanty S., 2012, ApJ, 746, 43

Ricker G. R., et al., 2014, Journal of Astronomical Telescopes, Instruments and Systems, 1, 014003

Ridden-Harper A. R., Snellen I. A. G., Keller C. U., Mollière P., 2019, A&A, 628, A70

Roberts S., McQuillan A., Reece S., Aigrain S., 2013, MNRAS, 435, 3639

Robitaille T. P., et al., 2013, A&A, 558

Russell H. N., 1923, Scientific America, 128, 78

Rutten R. G. M., Schrijver C. J., Lemmens A. F. P., Zwaan C., 1991, A&A, 252, 203

Saar S. H., Cuntz M., Shkolnik E., 2004, Symposium - International Astronomical Union, 219, 355–366

Saar S. H., Cuntz M., Kashyap V. L., Hall J. C., 2007, Proceedings of the International Astronomical Union, 3, 79

Salz M., et al., 2018, A&A, 620

Sanchis-Ojeda R., Rappaport S., Winn J. N., Kotson M. C., Levine A., Mellah I. E., 2014, ApJ, 787, 47

Sanchis-Ojeda R., et al., 2015, ApJ, 812

Santos A. R., Cunha M. S., Avelino P. P., Chaplin W. J., Campante T. L., 2016, MNRAS, 461, 224

Sanz-Forcada J., Ribas I., Micela G., Pollock A. M., García-Álvarez D., Solano E., Eiroa C., 2010, A&A, 511

Sanz-Forcada J., et al., 2011, A&A, 6, 1

Scandariato G., et al., 2013, A&A, 552

Scargle J. D., 1982, ApJ, 263, 835

Scharf C. A., 2010, ApJ, 722, 1547

Schlaufman K. C., Lin D. N. C., Ida S., 2010, ApJ, 724, L53

Schlichting H. E., Sari R., Yalinewich A., 2015, Icarus, 247, 81

Schölkopf B., Smola A., Müller K. R., 1997, in Proceedings of the International Conference on Artificial Neural Networks. Springer, Berlin, Heidelberg, pp 583–588, doi:10.1007/BFb0020217

Scholz F. W., Stephens M. A., 1987, Journal of the American Statistical Association, 82, 918

Schrijver C. J., Dobson A. K., Radick R. R., 1989, ApJ, 341, 1035

Schrijver C. J., Dobson A. K., Radick R. R., 1992, A&A, 258, 432

Schröder C., Reiners a., Schmitt J. H. M. M., 2009, A&A, 1107, 1099

Schröder K.-P., Mittag M., Pérez Martínez M. I., Cuntz M., Schmitt J. H. M. M., 2012, A&A, 540, A130

Schwartz J. C., Kashner Z., Jovmir D., Cowan N. B., 2017, ApJ, 850, 154

Schwarz G., 1978, The Annals of Statistics, 6, 461

Seager S., Sasselov D. D., 2000, ApJ, 537, 916

See V., Jardine M., Fares R., Donati J. F., Moutou C., 2015, MNRAS, 450, 4323

Seidel J. V., et al., 2019, A&A, 623

Sekiguchi M., Fukugita M., 2000, AJ, 120, 1072

Sheather S. J., 2009, A Modern Approach to Regression with R, 1 edn. Springer, New York, NY, doi:10.1007/978-0-387-09608-7

Sheets H. A., Deming D., 2017, AJ, 154, 160

Shizgal B. D., Arkos G. G., 1996, Reviews of Geophysics, 34, 483

Shkolnik E., Walker G. A. H., Bohlender D. A., 2003, ApJ, 597, 1092

Shkolnik E., Walker G. A. H., Bohlender D. a., Gu P., Kurster M., 2005, ApJ, 622, 1075

Shkolnik E., Bohlender D. A., Walker G. A. H., Collier Cameron A., 2008, ApJ, 676, 628

Sing D. K., et al., 2016, Nature, 529, 59

Singh K., 1981, The Annals of Statistics, 9, 1187

Sinukoff E., et al., 2016, The Astrophysical Journal, 827, 78

Skumanich A., 1972, ApJ, 171, 565

Smalley, B. et al., 2012, A&A, 547, A61

Smith, A. M. S. et al., 2014, A&A, 570, A64

Smith A. M. S., et al., 2012, AJ, 143, 81

Soderblom D. R., 2010, ARA&A, 48, 581

Soderblom D. R., Duncan D. K., Johnson D. R. H., 1991, ApJ, 375, 722

Sousa S. G., et al., 2019, MNRAS, 485, 3981

Southworth J., 2011, MNRAS, 417, 2166

Southworth J., Evans D. F., 2016, MNRAS, 463, 37

Southworth J., et al., 2009, MNRAS, 396, 1023

Spake J. J., et al., 2018, Nature, 557, 68

Spangler C., Sargent a. I., Silverstone M. D., Becklin E. E., Zuckerman B., 2001, ApJ, 555, 932

Spearman C., 1904, American Journal of Psychology, 15, 72

Staab D., 2018, PhD thesis, The Open University

Staab D., et al., 2017, MNRAS, 466, 738

Staab D., et al., 2019, Nature Astronomy, 4, 399

Steenbeck M., Krause F., 1966, Zeitschrift Naturforschung Teil A, 21, 1285

Stellingwerf R. F., 1978, ApJ, 224, 953

Stephens M. a., 1974, Journal of the American Statistical Association, 69, 730

Stepien K., 1995, MNRAS, 274, 1019

Stone M., 1977, Journal of the Royal Statistical Society, 39, 44

Strassmeier K. G., Washuettl a., Granzer T., Scheck M., Weber M., 2000, A&AS, 142, 275

Strugarek A., 2018, in Deeg H. J., Belmonte J. A., eds, Models of Star-Planet Magnetic Interaction, The Handbook of Exoplanets, 1 edn, Springer International Publishing, pp 1833–1855 (arXiv:1704.03254), doi:10.1007/978-3-319-55333-7_25

Strugarek A., Brun A. S., Matt S. P., Réville V., 2014, ApJ, 795, 86

Strugarek A., Bolmont E., Mathis S., Brun A. S., Réville V., Gallet F., Charbonnel C., 2017, ApJ, 847, L16

Strugarek A., Brun A. S., Donati J.-F., Moutou C., Réville V., 2019, ApJ, 881, 136

Student 1908, Biometrika, 6, 1

Sudarsky D., Burrows A., Pinto P., 2000, ApJ, 538, 885

Suveges M., 2012, in Proceedings of the Seventh Conference on Astronomical Data Analysis. p. 16, http://ada7.cosmostat.org/ADA7_proceeding_MSuveges2.pdf

Szabó G. M., Kálmán S., 2019, MNRAS, 485, L116

Szabó G. M., Kiss L. L., 2011, ApJ, 727, L2

Terquem C., 2014, MNRAS, 444, 1738

Teyssandier J., Lai D., Vick M., 2019, MNRAS, 486, 2265

Theil H., 1961, Revue économique, 14, 163

Tian F., Toon O. B., Pavlov A. A., De Sterck H., 2005, ApJ, 621, 1049

Tinetti G., et al., 2016, in Proceedings of Space Telescopes Instrumentation 2016: Optical, Infrared, and Millimeter Wave. SPIE, p. 99041X, doi:10.1117/12.2232370

Tobias S. M., 1997, A&A, 322, 1007

Torres G., et al., 2005, ApJ, 626, 523

Triaud A. H. M. J., et al., 2010, A&A, 524, 25

Trilling D. E., Benz W., Guillot T., Lunine J. I., Hubbard W. B., Burrows A., 1998, ApJ, 500, 428

Trilling D. E., Lunine J. I., Benz W., 2002, A&A, 394, 241

Tu L., Johnstone C. P., Güdel M., Lammer H., 2015, A&A, 577

Turner J. D., et al., 2020, ApJ, 888, L13

Udry S., et al., 2003, A&A, 407, 679

Valencia D., Ikoma M., Guillot T., Nettelmann N., 2010, A&A, 516

Valenti J. A., Fischer D. a., 2005, ApJS, 159, 141

Valsecchi F., Rasio F. A., Steffen J. H., 2014, ApJ, 793, L1

Van Dokkum P. G., 2001, PASP, 113, 1420

Van Eylen V., et al., 2016, AJ, 152, 143

Van Eylen V., Agentoft C., Lundkvist M. S., Kjeldsen H., Owen J. E., Fulton B. J., Petigura E., Snellen I., 2018, MNRAS, 479, 4786

VanderPlas J. T., 2018, ApJS, 236, 16

Vanderburg A., et al., 2015a, Nature, 526, 546

Vanderburg A., et al., 2015b, ApJ, 800

Vanderburg A., et al., 2015c, ApJ, 800, 59

Vanderburg A., et al., 2016, ApJ, 829, L9

Vapnik V., Chervonenkis A., 1964, Automation and Remote Control, 25

Vaughan A. H., Preston G. W., 2002, PASP, 92, 385

Vaughan A. H., Preston G. W., Wilson O. C., 1978, PASP, 90, 267

Vidal-Madjar A., Lecavelier des Etangs A., Désert J.-M., Ballester G. E., Ferlet R., Hébrard G., Mayor M., 2003, Nature, 422, 143

Vidal-Madjar A., et al., 2004, ApJ, 604, L69

Vidotto A. A., 2017, in EPJ Web of Conferences. (arXiv:1705.05561), doi:10.1051/epjconf/201716005011

Vidotto A. A., Cleary A., 2020, MNRAS, 494, 2417

Vidotto A. A., Fares R., Jardine M., Moutou C., Donati J. F., 2015, MNRAS, 449, 4117

Vines J. I., et al., 2019, MNRAS, 489, 4125

Voges W., et al., 1999, A&A, 349, 389

Walkowicz L. M., Hawley S. L., 2009, AJ, 137, 3297

Wang L., Dai F., 2019, ApJ, 873, L1

Weiss L. M., et al., 2017, AJ, 153, 265

Welsh B. Y., Lallement R., Vergely J.-L., Raimond S., 2010, A&A, 510, A54

West, R. G. et al., 2016, A&A, 585, A126

Wheatley P. J., et al., 2018, MNRAS, 475, 4476

Wilson O. C., 1963, ApJ, 138, 832

Wilson O. C., 1968, ApJ, 153, 221

Wilson O. C., Vainu Bappu M. K., 1957, ApJ, 125, 661

Winn J. N., Fabrycky D., Albrecht S., Johnson J. A., 2010, ApJ, 718, 2

Winn J. N., et al., 2017, AJ, 154, 60

Winn J. N., Sanchis-Ojeda R., Rappaport S., 2018, New Astronomy Reviews, 83, 37

Wisłocka A. M., Kovačević A. B., Balbi A., 2019, A&A, 624, A71

Wright J. T., 2004, AJ, 128, 1273

Wright J. T., 2005a, PASP, 117, 657

Wright J. T., 2005b, AJ, 129, 1776

Wright J. T., Miller B. P., 2015, Proceedings of the International Astronomical Union, 11, 357

Wright J. T., Marcy G. W., Butler R. P., Vogt S. S., 2004, ApJ, 152, 261

Wright J. T., Marcy G. W., Howard A. W., Johnson J. A., Morton T. D., Fischer D. A., 2012, ApJ, 753, 160

Wyman K., Redfield S., 2013, ApJ, 773

Xu W., Hou Y., Hung Y. S., Zou Y., 2013, Signal Processing, 93, 261

Yan D., Guo J., 2019, ApJ, 880, 90

Yan F., Henning T., 2018, Nature Astronomy, 2, 714

Yelle R. V., 2004, Icarus, 170, 167

Youngblood A., et al., 2016, ApJ, 824, 101

Zahn J. P., 2008, EAS Publication Series, 29, 67

Zhang M., Knutson H., Kataria T., Burrows A., Fortney J., 2018, AAS, 231, 211.06

Zhao S., Jiang-hui J., Yao D., 2018, Chinese Astronomy and Astrophysics, 42, 101

Zucker S., Mazeh T., 2002, ApJ, 568, L113

Zuckerman B., Song I., 2004, ARA&A, 42, 685